

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

СНЫТКИНА Дарья Алексеевна

Выпускная квалификационная работа

**Применение методов машинного обучения
при оценке и картографировании
природных ресурсов**

Уровень образования: магистратура

Направление 05.04.03 «Картография и геоинформатика»

Основная образовательная программа ВМ.5523. ««Геоинформационное
картографирование»

Научный руководитель: доцент, к.г.н., Кафедра Картографии и Геоинформатики,

Лазебник Ольга Анатольевна

доцент, к.г.н., Кафедра Картографии и Геоинформатики,

Капралов Евгений Геннадьевич

Рецензент: доцент, ФГАОУВО «Санкт-Петербургский политехнический

университет Петра Великого»,

Кикин Павел Михайлович

Санкт-Петербург

2020 г.

Содержание

Введение.....	2
Глава 1. Машинное обучение в решении задач использования природных ресурсов....	5
1.1. Искусственный интеллект и алгоритмы машинного обучения.....	5
1.2. Цели, задачи и функции машинного обучения.....	7
1.3. Модели машинного обучения в задачах классификации	11
1.4. Применение МО в задачах оценки природных ресурсов.....	14
1.5. Язык программирования Python и МО.....	18
Глава 2. Природные ресурсы как объект исследования и пространственной	
оценки.....	20
2.1. Виды природных ресурсов и особенности их исследования и оценки.....	20
2.2. Лесные ресурсы и лесное хозяйство: общая характеристика.....	22
2.3. Оценка лесных ресурсов.....	28
2.4. Зарубежный научный и практический опыт по оценке лесов с применением	
МО.....	31
2.5. Становление применения МО в российском лесном секторе.....	33
Глава 3. Изученность, использование, мониторинг бореальных лесов Северо-Запада	
России.....	36
3.1. Общая характеристика лесов.....	36
3.2. Экономическая значимость лесов Северо-Запада России.....	38
3.3. Действующие системы учета лесов Северо-Запада России.....	46
Глава 4. Создание модели классификации бореальных лесов с деградациями.....	49
4.1. Цели и задачи исследования.....	49
4.2. Выбор алгоритмов.....	50
4.3. Формирование набора исходных пространственных данных.....	50
4.4. Подготовка набора данных.....	53
4.4.1. Создание обучающей выборки для SAM и классификация.....	58
4.4.2. Создание обучающей выборки для Random Forest.....	64
4.4.3. Создание модели Random Forest и классификация.....	65
4.5. Оценка работоспособности модели МО.....	66
4.6. Обсуждение выявления повреждений древостоя методами МО.....	69
Заключение.....	72
Список использованных источников.....	74
Приложение. Модель RF (по национальному парку «Паанаярви»).....	85

Введение

Вовлеченность данных дистанционного зондирования Земли (ДЗЗ) в исследование и использование природных ресурсов постоянно растет, и в рамках этого процесса все более расширяется применение машинного обучения (МО). Особенно перспективными и эффективными эти технологии могут стать в отношении лесов - одних из самых высокопродуктивных и легкоранимых наземных экосистем, выполняющих средообразующие, защитные, эстетические функции. Злоупотребление сырьевой функцией лесов на протяжении долгого времени создало проблему обезлесения, усугубляемую неправильным упрощенным хозяйством, пожарами, болезнями леса, погодными условиями. Одновременно сохранить функции леса, а также увеличить выход деловой древесины позволяет рациональное лесное хозяйство, в основе которого лежит оперативный учет леса и его ресурсов. Частным примером успешного ведения лесного хозяйства является скандинавская модель. Внедрение именно такой модели инициировано органами государственной власти, природоохранными организациями и лесопромышленниками на Северо-Западе России, в историческом районе лесозаготовок. В связи с необходимостью интенсификации лесного хозяйства растёт спрос на технологии ускоренного получения пространственной информации о лесах и их состоянии. Если для организаций, производящих лесохозяйственные обследования традиционными методами, обширные территории представляют проблему, то при применении дистанционного зондирования и методов МО они, наоборот, являются преимуществом и означают наличие больших данных для обучения моделей. «Умная» обработка больших пространственных данных позволит минимизировать время- и финансово-затратные полевые обследования лесов.

Большинство предлагаемых исследователями методов обработки данных ДЗЗ с помощью МО носят локальный характер, однако в Скандинавии, Канаде, США отдельные алгоритмы включены в процедуры государственной инвентаризации. В лесной отрасли МО имеет существенный уклон в сторону прогнозирования пожаров и классификации пород. Вопрос обнаружения нарушений, влияющих экологическую и практическую ценность леса, освещается гораздо реже.

Возможности интеграции МО и данных ДЗЗ для выявления деградации лесного покрова изучали Крылов А.М. (Московская обл., Ленинградская обл., Вологодская обл.), Князева С.В. (Красноярский край), Карпачев А.П. (Орловская обл.), Abdi A. M. (Швеция).

Настоящее исследование является локальной трактовкой упомянутых выше разработанных технологий и открывает возможности выявления нарушений в бореальных

лесах Северо-Запада РФ на основе оперативных данных спутниковых наблюдений и алгоритмов МО.

Объектом исследования является нарушенный лесной покров в бореальных лесах Карелии на Северо-Западе России, **предметом** исследования – применение алгоритмов машинного обучения и спутниковых данных для обнаружения и оценки деградировавших древостоев в среде ГИС и в среде разработки Python.

Цель исследования состоит в применении алгоритмов машинного обучения к спутниковым данным для выявления нарушений в лесном покрове Северо-Запада России, а также в оценке работоспособности полученных моделей. Для достижения цели исследования требуется решить следующие **задачи**:

- ознакомиться с предшествующими исследованиями и выявить алгоритмы МО, наиболее часто применяемые для оценки природных ресурсов, и лесов в частности;
- изучить предметную область, основные причины деградации бореальных лесов Северо-Запада России и существующие методы идентификации таких участков;
- изучить алгоритмы МО на предмет требований, предъявляемых к данным и выделить оптимальные для анализа лесного покрова;
- сформировать набор пространственных данных о территории, включающие материалы спутниковой съёмки, тематические карты территории, материалы лесоустройства;
- с учётом особенностей алгоритмов МО создать обучающие выборки средствами геоинформационных систем и языка Python;
- осуществить программную реализацию модели классификации лесного покрова с деградацией древостоя;
- с помощью обученной модели получить информацию о состоянии лесного покрова, интерпретировать её в картографическом виде;
- сделать вывод о применимости исследованных моделей для классификации повреждений в лесах и об их действенности для оценки лесных ресурсов;
- разработать рекомендации по результатам исследования.

Научная новизна исследования заключается в наборе алгоритмов, которые были применены для решения задачи детектирования поврежденного древостоя, а также в том, что впервые были использованы материалы лесопатологических обследований для формирования контрольных данных о бореальных лесах Северо-Запада России.

В настоящем магистерском исследовании использованы методы интеллектуальной обработки данных, алгоритмы машинного обучения. В качестве информационной основы

использованы открытые мультиспектральные снимки с космических аппаратов Sentinel-2 и Landsat-8 за ряд дат 2016 - 2019 годов.

Теоретическая и практическая значимость исследования состоит в развитии подходов в применении методов МО совместно с данными ДЗЗ для обследования лесов и обнаружения их повреждений. Именно это может иметь большое значение для составления лесных карт, оценки биомассы с учетом пород, сопоставления имеющихся пород со спросом на них, управления лесом для повышения производительности отрасли и, конечно, для принятия мер по защите лесных экосистем.

Научно-исследовательская работа **состоит из** введения, четырех глав, заключения и списка литературы. Объем работы – 91 страница, 37 рисунков, 10 таблиц и 1 приложение. Список литературы содержит 144 наименования, в том числе Интернет-источники и иностранные публикации.

Работа выполнена на кафедре Картографии и геоинформатики СПбГУ.

Автор благодарен Евгению Геннадьевичу Капралову за влияние, оказанное при выборе предмета и объекта исследования, Ольге Анатольевне Лазебник за чуткое руководство, помощь в структуризации и грамотном изложении материалов работы, Разумовскому Владимиру Михайловичу за разъяснение подходов к оценке природных ресурсов, а также Орлову Василию за консультации по работе с лесоустроительными материалами.

Глава 1. Машинное обучение в решении задач использования природных ресурсов

В данной главе приводится краткое введение в предмет машинного обучения, предваряющее обсуждение использования его в контексте ГИС и в отношении пространственных данных. В обзоре мы опирались на работы А.С. Потапова, A.Thessen, Л.П. Коэлья, ресурсы сайта А.Дьяконова [1, 5, 9, 16].

1.1. Искусственный интеллект и алгоритмы машинного обучения

Вслед за автоматизацией и экспертными системами во все сферы деятельности человека планомерно и уверенно проникают искусственный интеллект и, в частности, машинное обучение.

Говоря про автоматизацию и разработку экспертных систем, А.С. Потапов подчеркивает, что разработчики устанавливали только *общие* правила, выведенные кем-то ранее, и закладывали их в код инструкций для механизмов. Такой подход наиболее приемлем, если спектр задач ограничен, и соответствующие действия для их выполнения четко определены. Но возьмем человека, обучающегося шахматам через наблюдение за игрой гроссмейстеров. Он запоминает ходы, но эти знания сможет применить, только в абсолютно идентичных ситуациях. Другой пример, для прогнозирования цены на автомобиле кроме пробега есть десятки комплектаций, разное техническое состояние, сезонность спроса и еще ряд неочевидных факторов, которые человек не учел бы в голове при всём желании и не заложил бы в формулу для расчета. Самим обучающимся интеллектом (будь то человек, или машина) должна выполняться дополнительная обработка информации с формированием общих правил на основе частных примеров [1].

В итоге, человеческий интеллект и стремление перекладывать трудовые обязанности заставили работать машины: находить закономерности в данных, учиться оптимизировать алгоритм (создавать формулу) и по нему получать ответ. Благодаря способности обрабатывать колоссальное количество информации современные машины стали находить даже такие закономерности, о которых люди не догадывались. Так родилось машинное обучение.

Учитывая растущий объем пространственных данных, механизмы, способные автоматически извлекать сложные структуры в таких пространственно-временных данных потребовались и в сфере наук о Земле. Для применения искусственного интеллекта для наблюдения Земли создан ряд инструментов, который уже доказал свою действенность в облегчении и ускорении многих видов работ.

В середине XX в. британский математик Алан Тьюринг сконструировал вычислительную машину, известную как «машина Тьюринга», способную воспроизводить когнитивные процессы. Для этого, считал Тьюринг, требуется разбить эти процессы на несколько отдельных этапов и представить алгоритмом. Так он заложил основу для того, что сегодня называется *искусственным интеллектом (ИИ)* [1].

Термины «*искусственный интеллект*» и «*автоматизация*» часто используются взаимозаменяемо. Они оба относятся к программному обеспечению, физическим роботам и другим машинам, способствующим повышению эффективности и действенность разных процессов. Однако уровни сложности двух систем очень разные.

Автоматизация заставляет что-то работать само, практически не взаимодействуя с людьми, следуя шаблонам и правилам для выполнения повторяющихся задач.

Искусственный интеллект помогает людям находить закономерности в неповторяющихся задачах, учиться на опыте, с помощью машинного обучения выбирать ответы. В отличие от автоматизации, ИИ не получает четко сформулированных команд и правил. Он призван освободить людей от выполнения утомительных рутинных задач (в области ГИС и ДЗЗ это, главным образом, автоматическое дешифрирование космических снимков) [2].

Машинное обучение, имитирующее человеческий интеллект, является областью информатики, связанной с созданием систем (рис.1.1). Иной областью компьютерных наук является *разработка программного обеспечения* (например, программирование на языке Python, C и пр.). Программирование можно назвать самым простым «обучением» компьютеров. МО подразумевает другой подход к обучению – автоматический поиск общего решения по частным примерам. Полнота информации, представленной в обучающей выборке служит основанием для отнесения обучения к одному из видов: с учителем, без учителя и с подкреплением [1].

Машинное обучение – дисциплина не новая. Своими успехами в последние годы она обязана практическому применению проверенных временем методов и идей из других областей знаний, в частности из математической статистики, линейной алгебры, теории вероятности и математической оптимизации. **Модели машинного обучения** строятся на основе статистики, теории вероятности и линейной алгебры, а затем оптимизируются. Успешным применениям машинного обучения посвящены ряд ресурсов (www.kaggle.com – платформа соревнований по анализу данных, www.forecsys.ru и др.): прикладная статистика всюду используется специалистами разных сфер. МО нашло широкое применение в совершенно разнообразных областях – в физике, бизнес-деятельности, медицине [3], строительстве [4], инженерной деятельности (в том числе экологической

инженерии) [5, 6]. МО и большие данные не обошли стороной и спорт. В 2019 г. в гонках королевской серии Ф1 появилось нововведение, разработанное при участии лидера облачных вычислений компании Amazon: в режиме реального времени сервер AWS получает миллионы элементов данных со 120 сенсоров болида, комбинирует их с 70-летними историческими записями о поведении машины на трассе, и с помощью модели МО в короткий срок дает прогноз, на основе которого тактики команд оптимизируют стратегию на оставшиеся круги [7].



Рис.1.1. «Инфраструктура» искусственного интеллекта [8].

Таким образом *машинное обучение* (МО) – это технология, способная извлечь из большого количества данных смысл путем нахождения шаблонов или закономерностей. Речь идет о таком объеме данных и таком формате их представления, которые человеческому интеллекту обработать невозможно. Давая определение МО также часто понимают, что это программное обеспечение, которое самостоятельно пишет программное обеспечение. Вместо применения предварительно созданной функции, МО приобретает опыт в процессе *обучения* и самостоятельно строит *модель* для применения в новых ситуациях.

1.2. Цели, задачи и функции машинного обучения

Цель машинного обучения, согласно Л.П. Коэльо и В. Ричарту, — научить программу решать задачу, предъявив ей несколько примеров [9]. Чем разнообразнее входные данные, тем проще машине найти закономерности и тем точнее результат. Для этого нужны данные, признаки и алгоритм (рис.1.2).

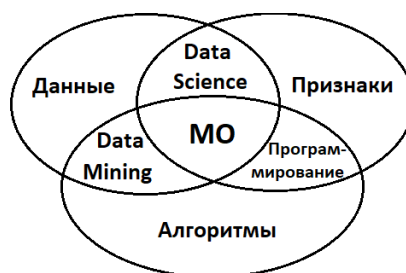


Рис. 1.2. Технология МО на стыке областей компьютерных наук, работающих с данными, алгоритмами и признаками [8].

Создание модели машинного обучения далеко не ограничивается настройкой обучающих алгоритмов. Л.П. Коэльо и В. Ричарт относят к этому:

- чтение и очистку данных (одна из самых трудоемких частей);
- изучение исходных данных и попытки понять их;
- конструирование признаков (feature engineering) – принятия решений о том, как лучше подать данные на вход алгоритма обучения;
- выбор подходящей модели (и алгоритма в ее основе);
- правильное измерение качества работы алгоритма (выбор метрики играет критическую роль) [9].

Система машинного обучения снабжается *входными данными* - числовыми, текстовыми, визуальными или аудиовизуальными. Также она имеет *выходные данные* – это может быть любое рациональное число (в таком случае говорят о регрессионных моделях), или целое число из ограниченного диапазона, представляющее категорию (класс) – для моделей классификации. Примеры: прогнозирование числа особей мышей по данным снимков Sentinel-2 и Landsat-8 [10], или отнесение пикселей или групп пикселей на снимке к одному из заранее определенных классов - урбанизированная территория или сельскохозяйственные угодья в задачах классификации земного покрова [11].

В зависимости от характера данных, задачи машинного обучения в целом делятся на три категории (табл. 1.1.):

- обучение без учителя (для задач кластеризации), контролируемое обучение с учителем (в задачах регрессии и классификации);
- полууправляемое обучение;
- обучение с подкреплением.

Табл.1.1. Типы машинного обучения (составлено по «Карте МО» из [8]).

Обучение	С учителем (контролируемое)		Без учителя	С подкреплением Ансамбли
Задачи	Классификация	Регрессия	Кластеризация Поиск правил	Компьютерное зрение Распознавание объектов
Алгоритмы	Наивный Байес, Деревья Решений (CART и C4.5), Логистическая Регрессия, К-ближайших соседей , Машины Опорных Векторов SVM	Линейная, Полиномиальная Регрессия		Random Forest , Gradient Boosting

Отдельные задачи можно решить разными методами или алгоритмами. От выбора метода зависит точность, скорость работы и размер готовой модели.

Самая популярная задача машинного обучения – **классификация**. Для нее всегда нужен «учитель» — размеченные данные с признаками и категориями, которые машина будет учиться определять по этим признакам.

Функциональная постановка задачи.

Предполагается, что существует неизвестная целевая зависимость — отображение $y^*: X \rightarrow Y$, значения которой известны только на объектах конечной обучающей выборки (X — множество описаний объектов, Y — множество допустимых ответов).

Тогда задача обучения с учителем (supervised learning) состоит в построении алгоритма $a: X \rightarrow Y$, который приближал бы неизвестную целевую зависимость как на элементах выборки, так и на всём множестве X [12].

В решении таких задач все чаще используют глубокое обучение и **нейронные сети**, как более продвинутые алгоритмы МО, способные совладать с более сложными данными, или **ансамблевые алгоритмы**, эффективность которых выше обычных методов обучения за счет их объединения в ансамбли и способности исправлять ошибки друг друга.

Сбор данных (X и Y) для обучения моделей может осуществляться как вручную — медленно, но по делу, так и автоматически. Качественный **набор данных — главное богатство разрабатываемой модели**. Из наборов данных выделяются признаки (они же свойства, характеристики или атрибуты), при этом все качественные показатели должны быть переведены в количественные. Для структурированных данных (в табличном представлении) признаки — это отдельные столбцы, характеризующие тот или иной показатель. Например, в задаче прогнозирования стоимости недвижимости набор данных может представлять собой таблицу, в строках которой собрана информация по уже проданным объектам, а столбцы содержат такие признаки, как площадь, местоположение (не координаты или адрес, а указание принадлежности к определенному району или расстояние от центра), уровень энергоэффективности здания, год постройки и целевой признак — цена [13]. Для неструктурированных данных, коими, в частности, являются космические снимки, экземпляры — это отдельные пиксели, а признаками являются в первую очередь их количественные параметры (например, значения интенсивности в разных каналах). В среде ГИС к пикселям раstra могут быть «подвязаны» прочие категориальные и количественные признаки, полученные из других наборов данных на территорию, охватываемую космоснимком [14].

Когда признаков много, модель работает медленно и неэффективно. Зачастую отбор правильных признаков занимает больше времени, чем непосредственно обучение.

Используемый для разработки набор данных при обучении разделяют как минимум на две части. Одна часть используется для разработки модели и называется тренировочным набором (условно, X_{train} и Y_{train}). Другая применяется для тестирования и носит, соответственно, названия тестового набора (X_{test} и Y_{test}). Разделение данных на обучающие и тестовые подмножества позволяет проверить, насколько хорошо модель, обученная на одних данных, может проводить классификации на новых.

Обученная модель, прошедшая валидацию, готова принимать на вход новые наборы данных и выдавать решение с некоторой точностью. В задачах машинного обучения для оценки качества моделей и сравнения различных алгоритмов используются различные метрики, а их выбор и анализ — неперенная часть работы.

В машинном обучении, согласно общей теории, различают *оценки качества* для задачи классификации и регрессии. Вводится функция потерь, характеризующая величину отклонения ответа модели от правильного ответа Y на произвольном объекте $x \in X$.

Типичный выбор функции потерь в задачах классификации $L(a, x, y) = [a(x) \neq y]$. Оценка задачи классификации часто значительно сложнее оценки регрессии [15].

Рассмотрим критерии качества в задачах классификации и важные при выборе метрики моменты.

Для описания метрик в терминах ошибок классификации используют матрицу несоответствий/ошибок (*confusion matrix*) – табличное сопоставление ответа алгоритма на объекте с истинной меткой класса на этом объекте. На рис. 1.3 показана такая матрица для решения задачи бинарной классификации. Два класса делятся на **положительный** (как правило, метка 1) и **отрицательный** (метка 0 или –1). Объекты, которые алгоритм относит к положительному классу, называются положительными (Positive); те из них, которые на самом деле принадлежат к этому классу – истинно положительными ИП (True Positive TP); остальные – ложно положительными ЛП (False Positive FP). Аналогичная терминология есть для отрицательного (Negative) класса: истинно отрицательные ИО (True Negative TP), и ложно отрицательные ЛО (False Negative FP). Ошибки классификации бывают двух видов: модель отнесла экземпляр к классу, к которому тот не принадлежит (*ошибка I-го рода*) или модель не отнесла экземпляр к классу, к которому он на самом деле принадлежит (*ошибка II-го рода*) [16].

Метка на объекте \ Ответ алгоритма	$a = 0$ negative	$a = 1$ positive
$y = 0$	ИО True Negative	ЛО False Positive
$y = 1$	ЛО False Negative	ИП True Positive

Рис. 1.3. Матрица ошибок [16].

Измерение качества классификатора начинается с принятия решения о том, что именно предстоит измерять. Самый простой и наивный способ – измерить среднее качество предсказания по всему тестовому набору (*accuracy*). Получится число между 0 (все предсказания неверны) и 1 (все предсказания правильны) [9]. Точность характеризует различающую способность того или иного признака.

Другие часто применяемые метрики - точность (*precision*), полнота (*recall*), а также F-мера (*F1-score*) - совместная оценка точности и полноты (рис. 1.4). F-мера позволяет получить более сбалансированную характеристику модели, чем три метрики, упомянутые выше [17]. А. Дьяконов призывает к аккуратному использованию термина «точность»: к сожалению, по-русски «accuracy» и «precision» переводят одинаково, но это разные метрики и их важно отличать.

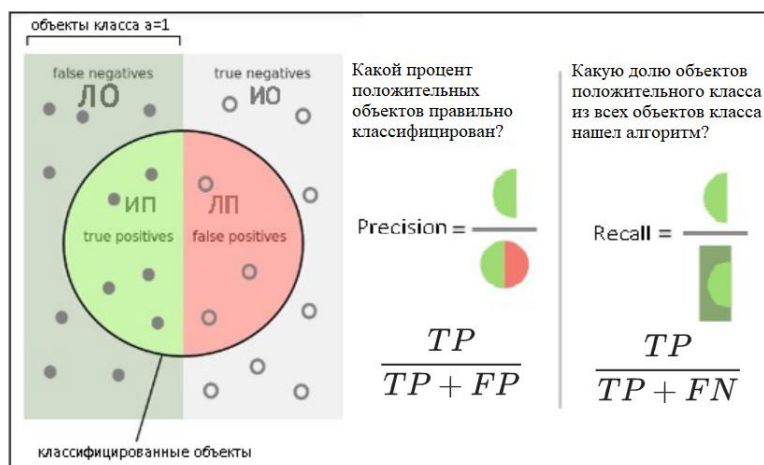


Рис. 1.4. Иллюстрация расчета точности и полноты классификатора [18]

В случае многоклассовой классификации нужно внимательно следить за метриками каждого из классов и следовать логике решения задачи, а не оптимизации метрики.

1.3. Модели машинного обучения в задачах классификации

В основе моделей машинного обучения лежат алгоритмы, часто эти понятия взаимозаменяемы.

Модели МО являются приближением процесса генерации данных; математически – это функции со всеми входящими в неё коэффициентами. Коэффициенты определяются через процесс *обучения* модели по *определённому алгоритму*. Непосредственный вид функции может быть задан заранее, тогда модели называются *параметрическими*; если функции формируются в ходе обучения («выучены»), их называют *непараметрическими*. Для последних обычно требуются большой объём обучающих данных для оценки основной функции. Классификаторы являются преимущественно непараметрическими моделями.

На сегодня известно множество алгоритмов машинного обучения, а также регулярно появляются новые.

Опираясь на работы [5, 14, 19 – 25], перечислим алгоритмы машинного обучения, чаще всего применяемые в задачах классификации.

Логистическая регрессия – это способ определения зависимости между переменными, одна из которых категориально зависима, а другие независимы. В параметрических моделях, построенных на основе логистической регрессии, применяется логистическая функция, которая, в свою очередь, является функцией линейной регрессии (алгоритм обычной линейной регрессии воссоздает зависимость между зависимой и независимой переменными как линейную, а затем полученный результат использует при прогнозировании). Практическое значение логистической регрессии заключается в том, что она является мощным статистическим методом предсказания событий, который включает в себя одну или несколько независимых переменных. Это востребовано в таких задачах, как прогноз прибыли от определённого товара; оценка вероятности землетрясения в конкретную дату.

Байесовские модели – наивные байесовские классификаторы относятся к семейству простых вероятностных классификаторов. Они берут начало из теоремы Байеса, которая применительно к данному случаю рассматривает функции как независимые (это называется наивным предположением). На практике используется для автоматической привязки статей новостной ленты к тематическим рубрикам; для выявления эмоциональной окраски текста или для распознавания лиц и других паттернов на изображениях [26].

Деревья Решений – деревья решений (англ. "Decision Trees") позволяют сформировать правила классификации в виде древовидной структуры. Решение задачи классификации заключается в определении значения категориального выходного атрибута на основании входных данных. Для этого сначала производится оценка степени

корреляции входных и выходных значений, после чего обнаруженные зависимости описываются в виде узлов дерева.

Zhou, B. et al. сообщают о положительных результатах классификации типов почв с применением деревьев решений и байесовской модели. Полученные изображения были сопоставлены с существующими почвенными картами, составленными на основе полевых исследований. Оценивая точность полученных почвенных карт, авторы сочли возможным применения данных моделей для картографирования классов почв на исследуемой территории [24].

В методе *k* ближайших соседей *k*-NN – для классификации каждого из объектов тестовой выборки необходимо последовательно выполнить следующие операции [27]:

- Вычислить расстояние до каждого из объектов обучающей выборки;
- Отобрать *k* объектов обучающей выборки, расстояние до которых минимально;
- Класс классифицируемого объекта — это класс, наиболее часто встречающийся среди *k* ближайших соседей.

Ученые многих стран используют именно *k*-NN как инструмент для проведения инвентаризации и картографирования земель лесного фонда. Так, в статье [25] отмечено, что выбор метода ближайших соседей *k*-NN был мотивирован широкой применимостью данного классификатора при создания карт лесного хозяйства Европы, США и Канады. При этом, как один из простейших алгоритмов классификации, *k*-NN признается неэффективным на реальных задачах [27].

Нейронные сети (или искусственные нейронные сети, ИНС – англ. Artificial Neural Networks, ANN) вдохновлены функциональностью человеческого мозга и представляют собой упрощенную модель структуры биологической нейронной сети, имитирующей сложные функции, такие как генерация шаблонов поведения, познание, обучение и принятие решений. Такие модели обычно используются для задач регрессии и классификации. ИНС могут применяться в любой ситуации, в которой существует взаимосвязь между предикторными и прогнозируемыми переменными независимо от сложности взаимосвязи. Они успешно применяются в сельском хозяйстве для управления растениеводческим хозяйством, выявления сорняков, болезней или иных характеристик [21], а также в управлении водными ресурсами [22]. Замай С.С. с коллегами разобрали связь ИНС с ГИС [28]. ИНС перерастает в глубокое обучение.

SVM (Support Vector Machine, Машина опорных векторов) – это бинарные классификаторы, в которых для классификации экземпляров данных (точек на плоскости или в пространстве) создается линейная разделяющая гиперплоскость. SVM используются для классификации, регрессии и кластеризации. В сельском хозяйстве они используются

для прогнозирования урожайности и качества сельскохозяйственных культур, [21], а Abolfazl Mollalo et.al. в своем исследовании по моделированию местообитаний конкретного вида мух признали его самым эффективным из рассмотренных и сумели составить карту [20].

Классификатор *Random Forest (RF)* – это ансамблевый алгоритм на основе большого количества бинарных деревьев решений. Он улучшает эффективность модели путём построения линейной комбинации более простых моделей. RF осуществляет случайный отбор наблюдений для «мелких» двоичных деревьев. Совокупность решений этого большого набора деревьев (леса) ведет к улучшению производительности метода. Окончательное решение достигается на основе большинства меток, полученных от отдельных деревьев.

Более сложные задачи, например, измерение уровня благосостояния живых организмов, требуют различных подходов, таких как множественные системы классификаторов в ансамблевом обучении или байесовские модели - вероятностные графические модели, где анализ проводится в контексте байесовского вывода [5, 20].

Задачи классификации могут помочь детектировать аномалии. Все вышеупомянутые алгоритмы были использованы в задачах классификации объектов окружающей среды. Способность Random Forest эффективно обрабатывать данные с большим числом признаков и классов, а также наличие методов оценивания значимости отдельных признаков видится нам доводом для использования его в исследовании.

1.4. Применение МО в задачах оценки природных ресурсов

Приложения МО, как аналитического инструмента, очень разнообразны. Некоторые, на первый взгляд, экзотичны и неочевидны.

Если говорить о пространственных данных в общем, то, например, основатель фирмы Mapistry Райн Янох называет несколько потенциальных взаимодействий МО с ГИС [29]. На основании социально-экономических данных, собранных в разных местах, можно спрогнозировать доход магазина, который только планируется открыть. Для этого сначала экономические и социальные данные необходимо объединить с доходами магазинов в среде ГИС. Таким образом, могут быть оценены финансовые ресурсы.

Другая проблема, с которой сталкиваются инженеры и общественные службы, и которую можно переложить в задачу МО, состоит в том, что для многих старых карт (отсканированные бумажные копии, файлы САПР) неизвестны системы координат/проекции. МО может использоваться для автоматического выравнивания

контуров старых карт по известным базовым картам или аэрофотоснимкам с использованием элементов компьютерного зрения.

Обращаясь к природным компонентам, надо сказать, что доступность открытых данных наблюдения Земли (ЕО) в рамках программ Copernicus и Landsat представляет собой беспрецедентный ресурс для многих приложений ЕО, начиная от мониторинга землепользования и земного покрова (LULC), мониторинга урожая и прогнозирования урожайности, заканчивая контролем стихийных бедствий [30].

Обсуждая области применения МО, нельзя обойти нефтяной комплекс. Нефть – это ресурс, определяющий мировую экономику. Разработанные в Orbital Insight алгоритмы компьютерного зрения измеряют тени, отбрасываемые стенками резервуаров с сырой нефтью на опустившиеся крышки, чтобы рассчитать объем оставшегося сырья (рис. 1.5). Это позволяет заинтересованным сторонам предугадать тенденции в запасах нефти [31].

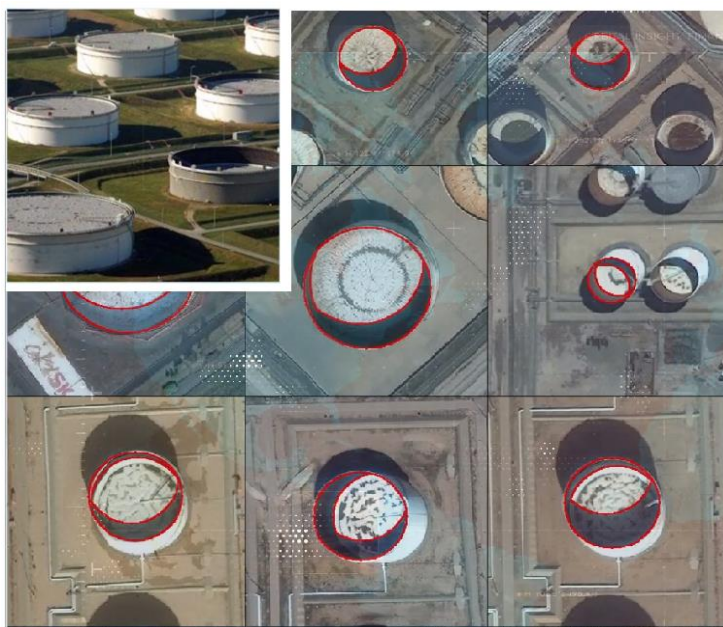


Рис.1.5. Нефтяные резервуары на космических снимках [31].

На Филиппинах тестируется решение на основе ИИ для мониторинга, описания и анализа устойчивости коралловых рифов, обеспечивающих среду обитания и кров примерно для 25% мировой морской флоры и фауны. Интеллектуальные подводные видеокамеры, оснащенные платформой прикладного интеллекта для обнаружения и фотофиксации рыб, ведут подсчет и классификацию морских организмов [32].

В сельском хозяйстве активно развивается точное земледелие (precision agriculture) — система управления продуктивностью посевов сельскохозяйственных угодий. Оно основано на комплексном использовании ГИС, данных ДЗЗ, систем глобального позиционирования и технологий оценки и прогнозирования урожайности, реализованных через алгоритмы МО. Особенностью данной области являются высокие требования к

регулярно обновляемой пространственной информации: данные должны обладать высоким пространственным разрешением и значительной повторяемостью. «Умная» обработка различных типов данных (климатические условия, виды почвы, рельеф и др.) позволяет как оценить потенциальные риски, так и предсказать, какой вид семян будет лучшим для посева в этих условиях, и сколько удобрения требуется. Это увеличит количество и качество выращиваемых культур.

Исследователи из Ирана пишут об интеграции методов машинного обучения в экологическое моделирование распространения вида песчаных мух [20]. Объектом исследования является тип мух *Phlebotomus papatasi**, главных переносчиков определенной инфекции, типичной для засушливых областей. Чтобы пресечь развитие инфекции и сохранить здоровье населения, важно максимально точно уметь спрогнозировать распространение ее переносчиков, что авторы стараются достичь, сравнивая несколько подходов моделирования – метод опорных векторов (SVM), случайный лес (Random Forest) и логистическую регрессию. В рамках ГИС сформирована географическая база данных, отобранных особей из разных частей региона; сопутствующие данные о среде, включая информацию по вегетации, топографию, параметры погоды, удаленность от главных рек; такие данные ДЗ как поверхность, индексированная значениями NDVI, и карты температуры поверхности земли LST. На выходе анализа (рис. 1.6) по приведенным данным получают матрицу вида есть/нет переносчик в данной точке пространства. Представление достоверной информации в виде интерактивной карты потенциально пригодных мест обитания вида мух может стать надежным подспорьем для лиц, принимающих решения по целенаправленным действиям.

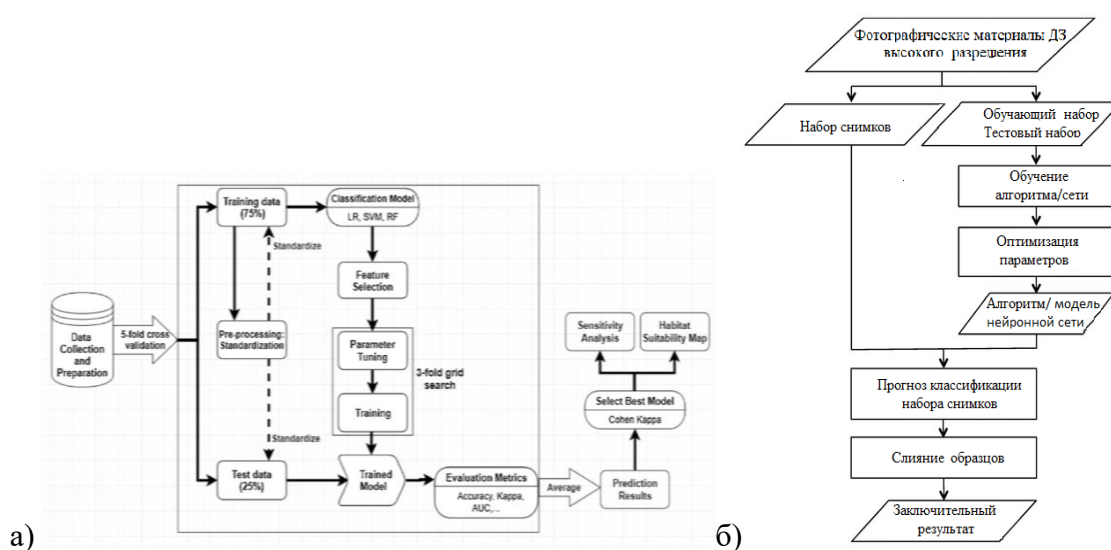


Рис.1.6. Функциональные схемы исследований по применению алгоритмов машинного к материалам ДЗ: а) частная и б) обобщенная [по 20].

С 2015 г. Институт мировых ресурсов (WRI) и “Orbital Insight” вместе работают над поиском новых приложений компьютерного зрения и машинного обучения, которые помогут “Global Forest Watch” в улучшении мониторинга лесов мира [33].

Проблемой некоторых регионов является замещение естественных лесов плантациями пальм в корыстных целях экономической выгоды. Кроны пальм и других деревьев неразличимы для привычных методов. В отличие от традиционных методов картографирования лесного покрова, основанных на степени «зелености» отдельных пикселей спутникового изображения, глубокое обучение (deep learning) рассматривает более широкий контекст изображения и может дифференцировать плантации на основе их цвета, размера, формы и рисунка. Способность Orbital Insight обрабатывать большие объемы данных позволила авторам названного проекта также использовать изображения с высоким разрешением от компании спутниковой съемки Planet. Чем выше разрешение снимков, тем легче идентифицировать структуры деревьев и дорог, более показательных для плантаций [33].

Planet – одна из компаний, специализирующихся на предоставлении пользователям геопространственной информации вместе с готовой аналитикой, необходимой для принятия обоснованных и своевременных решений. Общемировой тренд коммерческих фирм – продавать информацию, а не сырые данные. МО играет в этом одну из ведущих ролей [34].

Несколько лет назад компания запускала соревнование «Understanding The Amazon From Space» (Планета: понимание Амазонки из космоса) [35]. Сообществу Kaggle был брошен вызов: создать модель классификации спутниковых снимков бассейна Амазонки, чтобы лучше отслеживать и понимать причины обезлесения. Победившая модель представляет собой ансамбль из 11 нейросетей, способный выделить на снимке 17 типов покрытий, в том числе рубки и гари, выборочные рубки и ветровалы.

Применению МО в классификации покрова земли посвящено много научных работ [36-39]. Наиболее стабильные алгоритмы, доказавшие свою действенность, заложены в специализированные модули многозадачных ГИС: Classification Wizard в ArcGIS использует алгоритмы Максимального правдоподобия, Случайные деревья и Метод опорных векторов [40], SCP (Semi automated Classification Plug-in) в QGIS также использует алгоритм Максимального правдоподобия, Метод Спектрального Угла [41], в SAGA можно настроить k-NN. При этом важно понимать, что гибкость настройки встроенных алгоритмов ограничена функционалом конкретного ПО, поэтому при решении нестандартных задач приходится прибегать к написанию собственного кода, для чего может быть применен язык программирования Python.

1.5. Язык программирования Python и МО

Для конструирования моделей машинного обучения одним из самых удобных языков признается Python, подтверждением чему служит, например, число коммитов/репозиторий на GitHub. Обладая первоклассным набором библиотек машинного обучения с открытым исходным кодом, Python дает возможность сосредоточиться на решаемой задаче и в то же время опробовать различные идеи [9].

Изыскательская природа процесса поиска подхода к машинному обучению делает применение Python чрезвычайно уместным. Интерпретируемый высокоуровневый Python идеален для опробования разных вариантов. К тому же, он динамичен, хотя и не быстр. На Python численные расчеты можно перепоручить низкоуровневым расширениям, написанным, например, на C. Библиотеки можно найти на (<http://scipy.org/Download>).

Большинство современных алгоритмов оперирует высоко оптимизированными многомерными массивами – за них отвечает библиотека NumPy. А SciPy на базе этих массивов реализует быстрые численные алгоритмы: сложно найти в современных учебниках численный алгоритм, который не поддерживался бы в SciPy. Это относится к операциям над матрицами, линейной алгебре, оптимизации, кластеризации, пространственным операциям, преобразованию Фурье. Типовые алгоритмы разбиты на группы (packages), из которых наибольший интерес представляют пакеты `scipy.stats`, `scipy.spatial`, `scipy.interpolate`, и `scipy.cluster` [42].

Matplotlib (<http://matplotlib.org/>) - удобная и функционально развитая библиотека для построения высококачественных визуализаций на Python.

Выше указаны ссылки на вспомогательные библиотеки. Помимо них разработчиками было укомплектовано несколько специализированных библиотек, содержащих все необходимые функции для построения моделей МО: **scikit-learn**, **keras**, **eo-learn**.

В качестве моста, связывающего наблюдения за поверхностью Земли с разработкой на Python, выступает набирающая популярность библиотека **eo-learn** [43]. Это Python-библиотека с открытым исходным кодом соединяет дистанционное зондирование (Earth Observation) с экосистемой библиотек для машинного обучения в Python. Библиотека использует примитивы из библиотек `numpy` и `shapely` для хранения и манипуляции данными со спутников [44].

Изучение большого числа публикаций, освещающих основные вопросы и проблемы применения МО, показывает, что всё больше и больше людей применяют его в практической работе и научных исследованиях, тем самым закрепляя за МО звание

передовой технологии. Количество алгоритмов (моделей) МО растёт, и вместе с этим повышается точность результатов работ моделей. Обратим внимание на то, что, несмотря на значительную разработанность применения МО в таких областях как прогнозирование финансового рынка, медицина и др., технологии машинного обучения медленнее, но так же уверенно входят во многие направления наук о Земле, в частности, в мониторинг и анализ состояния природной среды. Есть основания полагать, что в ближайшем будущем большинство аналитических задач будут выполняться алгоритмами МО. Многие задачи в интересах картографирования и анализа, например, растительного покрова, с применением данных дистанционного зондирования, которые прежде казались трудно- или невозможно- решаемыми, могут быть реализованы с помощью алгоритмов МО.

Глава 2. Природные ресурсы как объект исследования и пространственной оценки

2.1. Виды природных ресурсов и особенности их исследования и оценки

В Российской Федерации понятие «природные ресурсы» законодательно закреплено статьей 1 Федерального закона «Об охране окружающей среды» [45]. Установлено: «*природные ресурсы* - это компоненты природной среды, которые используются и/или могут быть использованы при осуществлении хозяйственной деятельности». Исчерпывающее толкование дает В.М. Разумовский: «*природные ресурсы* – компоненты природы, необходимые для жизнедеятельности общества, эксплуатация которых в настоящее время и в обозримом будущем технически возможна, экономически целесообразна и экологически допустима». Однако еще больше расширила понятие «ресурс» парадигма устойчивого развития: от упора на исключительно составляющие биосферы, являющиеся рыночным товаром – древесина, рыба и пр., к ресурсам без рыночной цены – чистый воздух и вода, природное разнообразие, развитие человеческих ресурсов и др. [46].



Рис. 2.1. Экологическая классификация природных ресурсов [47].

Имеется несколько классификаций природных ресурсов и способов охарактеризовать лес. Природная классификация исходит из того, что природные ресурсы являются компонентами биосферы, экономическая — из направления и формы использования ресурсов. Широко используется важная с точки зрения охраны природы классификация природных ресурсов по их истощаемости или экологическая (рис.2.1). В соответствии с ней природные ресурсы разделяют на *исчерпаемые* и *неисчерпаемые*. К *возобновляемым* ресурсам, способным к самовосстановлению через сроки, соизмеримое с хозяйственной деятельностью человека, относится большинство видов растительного и животного мира, в том числе лес [48].

Деятельность человека в сфере его отношений с природой включает: *природоведение, природопользование и природообустройство*. Попытки изменить природную среду на благо человека привели к появлению недопустимо частых негативных последствий в ее состоянии, то есть к современному экологическому кризису. Ответом кризису становится постепенное внедрение рационального природопользования, в основе которого лежат капитальные знания о природном комплексе и предыдущий опыт решения административных и финансовых вопросов. Государства мира находятся на разных стадиях принятия необходимости мер и их имплементации [49].

В структуре биологических ресурсов одним из важнейших, если не самым важным и стратегическим ресурсом, является **лес**.

Лес – сложнейшая многокомпонентная система и стратегический ресурс многих государств (и РФ), поэтому он является предметом интереса государственных и научных структур. Леса относятся к возобновляемым источникам сырья, что значит, что скорость восстановления древесных природных ресурсов, формирующих биомассу, сравнима со скоростью их расходования. Это определение справедливо только тогда, когда, сводя леса с определенной энергетической или вещественной целью, не забывают удостовериться в том, что скорость восстановления или замещения другой породой действительно выше. Для выполнения этого условия государства вводят четкую регламентацию и нормы лесного права.

Будучи огромным углеродным пулом [50, 51, 52], лесные системы - критически важный компонент глобального углеродного цикла. Точные **оценки** надземной биомассы лесов планеты (английский термин AGB – aboveground biomass) помогают заглянуть в его механизм и помочь подсчитать антропогенные выбросы для более эффективного регулирования [53].

Существенна роль лесов в создании **экологически и эстетически** комфортной среды обитания и меры для ее реализации включают анализ ряда параметров древесных насаждений.

Леса лучше других типов растительности выполняют **противоэрозионные функции**, оказывают положительное влияние не только на территории непосредственно под лесом, но и на прилегающие к нему. Они выполняют роль **экотонов** – буферных участков, представляющих собой барьеры для мигрирующих в ландшафте веществ [54, 55]. Сохранение лесных массивов, ключ к которому их оперативный **учет**, должно помочь в достижении экологического равновесия.

Лесные объекты подлежат лесоустройству и охране. Обсуждая лесные ландшафты запада и востока Европы, Р. Angelstam et. al. [56] показывают, что лесопользование и

сохранение биологического разнообразия плохо совместимы с точки зрения лесоустройства.

Обезлесение, зафиксированное с начала промышленной революции, заставило правительства и других стран мира признать глобальную роль лесов и лесного хозяйства и обратить внимание на проблему обезлесения, но действия долго откладывались [57]. Основной проблемой в обеспечении функционирования зеленой инфраструктуры является слабое межотраслевое взаимодействие [56]. Ключами к оптимальному функционированию зеленой инфраструктуры авторы видят в получении новых знаний, поддержке исследований и обучения в отношении лесов, а также в усилении межотраслевого взаимодействия.

Нам представляется, что некоторым вкладом в дело сохранения лесов может стать решение задачи применения машинного обучения в исследовании и картографировании лесных массивов с применением данных дистанционного зондирования на примере одного из регионов России.

2.2. Лесные ресурсы и лесное хозяйство: общая характеристика

«Невозможно представить развитие цивилизации и счастье человека на земле, лишенной лесов» Н.С. Нестеров

Лесной Кодекс Российской Федерации [58], самого «лесного» государства [59], рассматривает лес как дуалистический объект – и как экосистему, и как природный ресурс. Определим место и значимость леса и его ресурсов в природном и политико-экономическом каркасах.

Использование, охрана, защита, воспроизводство лесов в России осуществляются исходя из понятия о лесе как об экологической системе или как о природном ресурсе на основе лесного законодательства. Леса располагаются на землях лесного фонда и землях иных категорий, классификация земель лесного фонда приводится ниже (табл. 2.1). Земли лесного фонда состоят из лесничеств и лесопарков.

Табл. 2.1. Структура Земель лесного фонда [60].

Земли Лесного Фонда					
Лесные					Нелесные
Покрытые лесом	Несомкнувшиеся лесные культуры	Лесные плантации и питомники	Естественные редины	Не покрытые лесом	
Молодняки: полнота > 0,4; Иные группы возраста: полнота > 0,3	Плантации пород для ускоренного выращивания древостоев с целью получения целевых сортиментов или древесной массы для хим. переработки.	Земли для выращивания лесопосадочного материала	Древостои: полнота 0,1 - 0,2	Древесно-кустарниковой растительности нет, общие показатели полноты, сомкнутости или количества экземпляров древесных растений недостаточны.	

участки, занятые кустарниками	По таксационным показателям и состоянию относят к покрытым лесом землям или к несомкнувшимся лесным культурам.				
вырубки, гари (естественно возобновляющиеся участки): количество и состояние экземпляров естественного возобновления или сохраненного при рубке леса подроста соответствует нормативному или превышает его;		плантации новогодних елей	Гари	погибшие насаждения - древостой, усохшие на корню / полностью прекратившие вегетацию в результате:	
			массового повреждения фитовредителями	стихийного воздействия (ветровала, бурелома, снеголома)	выбросов в атмосферу вредных веществ;
			Вырубки (древостой полностью вырублен в порядке главного пользования лесом или в результате сплошных санитарных рубок)		
			прогалины и пустыри		
		молодняк: общая полнота или сомкнутость < 0.3			

На глобальном уровне протяженные леса России участвуют в круговороте мировых запасов углерода, они играют ключевую роль в динамике парниковых газов в пределах бореальной зоны страны [61, 62]. Леса должны составлять особую заботу ответственных и заинтересованных лиц, поскольку только при **рациональном хозяйствовании** они могут эффективно депонировать излишки углерода и улучшать природную среду. Помимо этого, грамотная лесоэксплуатация обеспечивает повышение ресурсного потенциала, и тогда леса могут приносить доход вечно (в отличие от полезных ископаемых). В экономике России, значительные объемы древесины потребляются и на внутреннем рынке, и составляют существенную долю экспорта, беря на себя **роль стратегически важного ресурса**.

Согласно статье 67 Лесного Кодекса РФ на землях лесного фонда, а также на землях особо охраняемых природных территорий (ООПТ), на которых расположены леса, проводится обязательное **лесоустройство**. Основные отличительные черты лесов

лесного фонда и ООПТ, с точки зрения лесного дела, состоят в характере изъятия лесов из хозяйственного использования. Цели при организации и ведении хозяйства в лесах разных категорий иные, а, следовательно, и намечаемые мероприятия другие. Важным мероприятием при лесоустройстве считается *таксация леса*. При таксации лесов ООПТ желательно применять сплошной измерительно-перечислительный метод (согласно «Основным положениям о проведении лесоустройства национальных парков» от 7.07.1993 г.) [63]. Нам представляется, что данные лесоустройства и таксации лесов ООПТ могут стать ценным материалом в применении МО в исследовании и картографировании лесов.

Лесное хозяйство является важной отраслью **экономики** многих стран. На территорию России приходится 1/4 всех мировых запасов древесины. По данным за 2019 г. общая лесная площадь превысила 894 млн га, что составляет 45% всей площади страны [64]. При этом объем запасов составлял около 82 млрд м³ [65].

Примерная структура лесных насаждений, произрастающих на территории России, приведена на рис. 2.2.



Рис.2.2. Диаграмма породного состава лесов России (составлено по [66]).

Доля экономически эксплуатируемых лесных насаждений в России достигает 51 %. Остальные леса имеют «непромышленное» значение (экологическое, рекреационное). Экономически значимые лесные насаждения сконцентрированы в Республике Карелия, Архангельской и Амурской областях, Красноярском крае, Республике Коми. Здесь темпы вырубки значительно превышают темпы воспроизводства ресурсов. Вопрос о рационализации лесохозяйственной деятельности России стоит остро, хотя доля лесной промышленности в ВВП России составляет всего 2 %.

Идея *пользования лесом* является *доминантой лесного хозяйства* и лесного законодательства России. Государственное значение лесу было придано при Петре I,

который, стремясь к повсеместному наведению порядка, установил платный отпуск ресурса. Небрежное отношение к лесам России происходило от убеждения, что лесов в России много. Однако, лес хоть и возобновимый, но все же исчерпаемый ресурс. Важнейший критерий неистощительности - **соблюдение нормы пользования** (золотое правило устойчивости Г.Л. Гартига), т.е. связь вырубленной древесины с приростом. «Удар» по лесу нанес научно-технический прогресс: резко вырос масштаб вырубок на стройматериалы, как источника энергии, бумага стала товаром массового потребления, а механизация и снижение требований к качеству раздвинула зона экономически доступных лесов. В России середины XX в. промышленные лесозаготовки переместились из истощенных центральных районов в многолесные северные леса европейской России, в результате чего в Карелии показатели лесистость снизились на 20%, достигнув 48% территории. Масштабы трансформации и деградации лесных экосистем, особенно бореальных (северных и таежных), приобрели опасный размер.

К концу XX в. ситуация в России улучшилась благодаря проведению мероприятий по содействию естественному возобновлению, соблюдению правил порубок, внедрению новых технологий лесозаготовок и, в принципе, благодаря появлению других строительных и энергетических ресурсов [67, 68].

Управление лесами лесным хозяйством осуществляется Федеральным агентством лесного хозяйства России – Рослесхозом [69]. Его структурными подразделениями являются:

Рослесинфорг	Авиалесоохрана	Российский центр защиты леса	Рослесресурс	НИИ Лесного хозяйства, СПб
Государственная инвентаризация леса (ГИЛ), дистанционный мониторинг за использованием леса, разработка Лесных планов, лесоустройство, планирование и проектирование;	Контроль за достоверностью сведений о пожарной опасности;	Лесопатологический мониторинг (ЛМП);	Организация экономических цепочек реализации ресурсов леса и пр.;	Среди основных направлений деятельности – разработка теории и практических методов внедрения модели интенсивного лесного хозяйства и применения в нем матем. методов и информ. технологий.

В России действует Лесной Кодекс, а также есть ГОСТ Р 56695-2015 «Возобновляемые источники сырья. Лесные ресурсы», в котором собраны основные термины и понятия.

В Рослесхозе имеется три государственные информационные системы для ведения лесного хозяйства:

- Единая государственная информационная система учёта древесины и сделок с ней (ЛесЕГАИС) – для учёта древесины на основе представленной в нее информации;
- информационная система дистанционного мониторинга (ИСДМ-Рослесхоз) – для принятия управленческих решений по борьбе с лесными пожарами;
- система электронного документооборота (СЭД Рослесхоза) [70].

Сегодня утверждение про изобилие леса в России – не полная правда; пришло понимание важности комплексного подхода к оценке леса перед его эксплуатацией: краткосрочная экономическая выгода приведет к значительным затратам на восстановление равновесного состояния.

Интенсификация производства древесины угрожает биоразнообразию. Тем не менее, в лесном деле есть и другие *проблемы*. Динамика лесных экосистем продолжается, и важным ее фактором являются различные *естественные нарушения*: *пожары, заболевания, ураганные ветры* (см. табл. 2.1). Последствиями пожаров, катастроф и рубок являются макро-нарушения, (пространственный масштаб более 200 м², рис. 2.3) [46]. Это попадает в пространственное разрешение космических снимков некоммерческого сектора. Это значит, что данные ДЗЗ открытых сервисов Copernicus Open Access Hub, Sentinel Hub, LandViewer и др. доступны для изучения подобных нарушений. Для дистанционного мониторинга процессов повреждения отдельных деревьев, а также при площади нарушенного участка менее 200 м², нужны снимки более высокого разрешения, доступ к которым возможен только на коммерческой основе.



Рис. 2.3. Виды нарушений в лесных экосистемах. Классификация по масштабу [46]

В последние годы, и в 2019 г. особенно, лесные **пожары** и в России и за рубежом достигли катастрофического уровня [71-73]. Причины их возникновения самые разные: засуха, высокая температура и сильные ветра, поджоги и др. Существует мнение, что

поджогами пытаются скрыть незаконные вырубки леса. Общая площадь пожаров в России в 2019 г. составила больше 5 млн га, погибло более 60 тыс. га леса, а экономический ущерб превысил 14 млрд руб. [74]. Пожары также влияют на качество воздуха, на здоровье людей и животных, уничтожают их естественные места обитания. Несмотря на то, что пожары возникают и в естественных условиях, а в таких видах экосистем как сосновые и лиственные леса ими даже обуславливается структура (она представляет собой мозаику гарей разного возраста), виновник большинства возгораний – человек.

После пожаров необходимо измерять поврежденные участки, чтобы на основании данных о площади гарей планировать работы по восстановлению растительности. Даже после катастрофических пожаров вдоль водотоков могут сохраниться фрагменты древостоя (пожарные рефугиумы), которые важны для восстановления биоразнообразия, и которые, соответственно, важно выявлять и не допускать их рубки [46]. Там, где пожары проявляются как естественный фактор, отслеживание их масштабов важно для служб управления лесами.

Наиболее практичным способом мониторинга последствий пожаров и информационного обеспечения планирования восстановительных мер является использование данных ДЗЗ. Космические снимки позволяют быстро оценить последствия лесных пожаров и помогают правильно выстроить стратегию по восстановлению растительности. За несколько этапов обработки космоснимков можно найти пользовательскую комбинацию каналов, демонстрирующую гари наиболее наглядно, а затем с помощью инструментов ГИС произвести векторизацию поврежденных пожарами контуров и вычисление их площади. С использованием ГИС и спутниковых изображений до и после пожаров становится возможным вычисление нормализованного индекса гарей (NBR - Normalized Burn Ratio) и составление тематических карт [75].

После бесснежной и теплой зимы 2020 г. Гринпис России и Сообщество Природоохранных ГИС взяли за задачу создания слоя ландшафтных весенних пожаров по всей территории России [из личной рассылки]. Такие сообщества менее инерционны, чем ведомственные подразделения, и, тем не менее, они ищут помощь добровольцев в оцифровке, что говорит об отсутствии автоматизации в данной сфере.

Другой проблемой лесной промышленности являются **заболевания** деревьев. Действующая методика оценки состояния древесного фонда требует выезда на местность и произведения замеров доли ослабленных деревьев по категориям с указанием причин. Эти данные заносятся в акт лесопатологического обследования (ЛПО). Однако с расширением использования аэрокосмических методов и ГИС, стало возможно обследование здоровья деревьев по данным ДЗЗ. Заболевания растительности

отображается на снимке (в прямом смысле), поскольку изменение химического состава клеток заставляет листву поглощать и отражать электромагнитное излучение иной частоты, нежели здоровое растение, что и фиксируется на снимке. Комбинируя с каналами мультиспектрального снимка, можно вычислить такие индексы, как NDVI, SWVI и др. и детектировать территории с проблемным древостоем.

В случае если дерево гибнет, но остается стоять на корню (заболевание), лесная среда локально не сильно изменяется. Если же дерево падает - от старости или ураганного ветра, то изменения становятся заметнее, и на данном участке образуются сразу несколько более мелких экосистем (мертвая древесина на почве, ветровально-почвенные комплексы - бугор и западина). Разрыв в лесном пологе изменяет условия освещенности и, соответственно, параметры микроклимата (плотность полога леса – одна из ключевых характеристик, для создания моделей которых разработано много методических подходов). Новые комплексы могут стать местом обитания специфических организмов, в том числе и *недревесных ресурсов леса* – ягод или других видов растений. Важно понимать, что отмирание деревьев – естественный процесс, ведущий к образованию окон возобновления, которые в свою очередь нужны для роста светолюбивых видов. Через некоторое время на участках, подвергшихся воздействию сильных ветров, образуются сухостой.

Проведённое рассмотрение общих черт лесных ресурсов, а также основных проблем их сохранения, позволяет выявить возможные направления применения машинного обучения для распознавания естественных нарушений этих ресурсов.

2.3. Оценка лесных ресурсов

Главной функцией леса, как правило, считается *сырьевая*. Существует много методов количественной оценки древесины, учитывающих энергетические параметры отдельных пород, а также и других ресурсов леса: водных, рекреационных, охотничьих, рыбных, пастбищных и др. Чтобы ресурсы не истощались, необходимо соблюдать нормы пользования и грамотно закладывать это в планы по лесоэксплуатации с учетом всей информации. Не менее активно ведутся работы по оценке *углеродного цикла* лесного покрова, что связано, главным образом, с массовой обеспокоенностью изменением климата, вызванным антропогенным повышением атмосферных концентраций парниковых газов (в первую очередь углекислого газа) [76].

Оценка состояния лесных ресурсов является отдельным направлением лесной науки и практики. Она осуществляется в рамках государственного управления посредством *таксации* (от лат. *taxatio* – оценка) леса – комплекса мероприятий по учету

и всесторонней оценке, определения технических характеристик насаждений и их возраста, объема отдельных деревьев и, конечно, запаса древесины.

Оценка состояния лесов в мире проводится под эгидой Всемирной продовольственной организацией ООН (FAO), и РФ, как самая лесная держава, конечно, участвует в ней.

В России проводятся следующие оценки:

- оценка использования ресурсов леса населением для собственных нужд;
- оценка потребности и обеспеченности сырьем промышленности, перерабатывающей лесные ресурсы;
- оценка потенциала и фактического использования лесов с целью заготовки и сбора недревесных лесных ресурсов;
- оценка рекреационного потенциала и объемов фактического использования лесов в целях рекреационной деятельности;
- **оценка экологического потенциала**, потенциала средообразующих, водоохраных, защитных, санитарно-гигиенических и иных полезных функций лесов;
- оценка потенциала лесов для осуществления видов деятельности в сфере охотничьего хозяйства; оценка потенциала и фактического объема использования лесов для ведения сельского хозяйства;
- а также это могут быть **производные оценки** такие как: **ожидаемый доход** от реализации заготовленной древесины, планируемые затраты на проведение мероприятий (рубок ухода и др.), прибыль на единицу площади участка леса [77, 78, 79].

Единицами оценки являются лесные участки, на которые делятся все земли лесного фонда (и иные земли, на которых расположены леса). Лесной участок – земельный участок, границы которого определяются в соответствии с материалами лесоустройства и данными кадастрового учета. Лесное законодательство предусматривает обязательный кадастровый учет лесных участков. Работы по отводу и таксации лесосек в лесах РФ регламентируются Правилами заготовки древесины [80]. В отношении лесов, расположенных на землях лесного фонда, проводится ***государственная инвентаризация лесов***. Она осуществляется наземными и аэрокосмическими способами. Важно понимать, что при подборе данных ДЗЗ для использования в мониторинге состояния лесов, их инвентаризации и оценке необходимо опираться на оптические свойства лесных территорий, связанные с их сезонным развитием.

Результаты инвентаризации предоставляются в органы государственной власти субъектов РФ, уполномоченные в области лесных отношений. В настоящее время ведется активная работа над порталом о лесных ресурсах страны, запуск которого планируется в

2020 г. Благодаря наличию функционирующего портала станет возможным оперативно узнавать не только о лесных ресурсах, но и о лесах, гибнущих от пожаров и вредителей, а также о лесовосстановлении [81].

Инвентаризацией и картографированием лесов по данным ДЗЗ (космическим снимкам) по договорам Рослесхоза занимаются коммерческие компании. Компания «Совзонд» использовала снимки с RapidEye, и на их основе создала базовые электронные карты квартальной сети, выделяла покрытые лесом площади. Еще в 2009 г. по снимкам с высоким разрешением было начато определение категорий (лиственные, хвойные) и породного состава древостоя, выявление ветровалов, а также их картографирование [82].

Для наполнения платформ контроля лесной отрасли необходимы оперативные и точные данные. Это направление работ развивается по стране в целом. Рослесхоз при подготовке федерального информационного ресурса готов брать на вооружение наилучшие практики регионального уровня, успешно вовлекающие в свои работы данные ДЗЗ и геоинформационные технологии.

Так, РГИС Московской области имеет интерактивную карту, отображающую санитарные рубки, лесоводственный уход за культурами и уход в молодняках, месторасположения зон отдыха и лесные дороги [83]; на геопортале Калужской области можно найти информацию о категориях защищённости лесных массивов. Создатели данного геопортала строят планы на выявление и отображение незаконных вырубок [84].

По информации от портала forest.ru [85], одной из приоритетных задач ФБУ «Рослесозащита» на 2020 г. остается обновление информации о наличии земель, не занятых лесными насаждениями для их последующего лесовосстановления. Планирование лесовосстановительных работ подразумевает не только "заполнение пустот" лесопосадками, но также и прогнозирование лесовосстановления в будущем.

Центр по проблемам экологии и продуктивности лесов РАН, специализирующийся на разработках для лесного сектора, на своем сайте разместил информацию о методологии оценки таксационных характеристик по спутниковым данным, о методике проведения оценки ежегодных размеров пожарных эмиссий углерода (также по спутнику), о разработке программы для учета запасов валежа на площади. В математической модели программы используются методы учета валежа на линейных трансектах [86]. Программа имеет сдержанный интерфейс и, принимая на вход текстовый файл с четырьмя столбцами данных, вычисляет запасы биомассы, запасы валежа и запасы углерода (рис. 2.4).

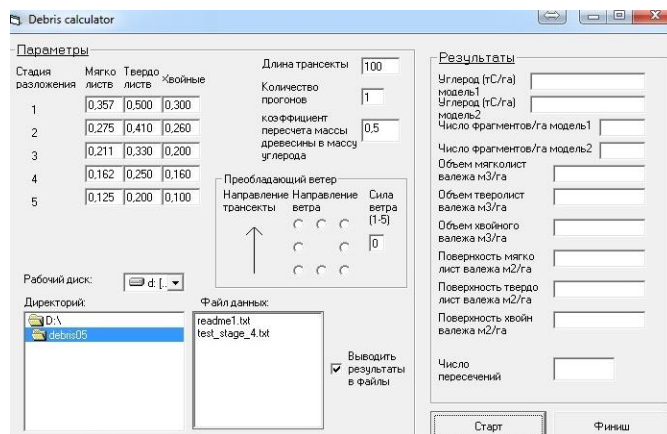


Рис. 2.4. Интерфейс программы для оценки запасы биомассы и углерода в валеже.

Таким образом, в лесном хозяйстве налицо тенденция к ускорению получения и использования пространственной информации о лесных массивах, а также стремление к получению прогнозируемых или планируемых событий на зеленых территориях.

Это подтверждает актуальность применения машинных алгоритмов прогнозирования и "умной" обработки больших данных, собранных с минимизацией натурных обследований.

2.4. Зарубежный научный и практический опыт по оценке лесов с применением МО

*«Над Канадой небо синее,
Меж берёз дожди косые...
Хоть похоже на Россию,
Только всё же — не Россия».*
А. Городницкий

Изучение научного и практического опыта применения машинного обучения в оценке лесов формирует представление о самых популярных зарубежных моделях и алгоритмах МО в решениях задач лесной отрасли и их преимуществах. Знакомство с публикациями по теме позволяет продвигаться в исследовании, опираясь на современные достижения, избежать возможных ошибок, а также разрешить возникающие в ходе работ вопросы.

Как было рассмотрено в Главе 1, машинное обучение – это набор четко определенных моделей, которые собирают конкретные данные и применяют конкретные алгоритмы для достижения ожидаемых результатов. Применительно к лесной отрасли, как показывают публикации, МО вовлекается неравномерно: преобладающее число

исследований посвящено прогнозированию пожарной опасности и классификации по породам.

Поскольку похожим видовым составом лесов, а также схожими климатическими условиями с Россией обладают скандинавские страны (Фенноскандия), а также Канада, стоит в первую очередь обратить внимание на успехи исследователей этих регионов в учете и управлении лесными ресурсами на основе ДДЗЗ, ГИС и искусственного интеллекта [87].

Первоочередным стимулом к изучению лесов для канадцев является потепление, наиболее заметное именно в циркумбореальных регионах и оказывающее влияние на чувствительные бореальные леса. С. Mahoney и соавторы [25] подчеркивают сложность картографирования обширных труднодоступных территорий Севера. Карты рассматриваются авторами как форма отображения инвентаризационной информации, ценной для заинтересованных лиц: властей, представителей промышленности и частных лиц. Для создания карт непрерывного распределения высоты насаждения и сомкнутости крон (плотности полога леса - по российскому ГОСТу это плотность стояния деревьев в древостое, характеризующаяся числом деревьев на единице площади (1 га) с учетом их толщины) канадской тайги они задействуют данные с бортов космических аппаратов (КА), лидарные облака точек и незначительный объем от полевых исследований, извлекая информацию посредством k-NN. Высота насаждений важна в первую очередь для определения объема надземной биомассы (AGB). Результатом их исследования является механизм, который планируется развивать с привлечением данных PALSAR в модели Random Forest, и тем самым расширить для оценки и картографирования дополнительных переменных учёта лесного фонда.

Алгоритм k-NN является функцией пространственных данных, полученных из разных источников. Он в целом широко используется при инвентаризации и картографировании не только в Канаде, но и в США, сопоставимых с Россией по площади, занятой лесами, и в странах, определяющих тенденции развития государственных систем инвентаризации лесов в Европейском регионе (Швеция, Финляндия, Германия) [88, 89].

Определенный интерес представляет информация D. Stojanova и A. Kobler о формировании набора данных для классификации трех лесных территорий Словении по пожароопасной ситуации [90]. Авторы использовали ряд ГИС-слоев (медианная высота над у.м., тип экспозиции, дистанция до дорог, ж/д и др.), космические снимки MODIS и LANDSAT, облака точек (LiDAR), набор метеоданных ALADIN. Как говорилось ранее, подготовка машиночитаемых данных – часто ключевая часть работы с МО, и авторы

данного исследования приложили к этому много усилий. Они делают акцент на том, что важно соблюдать равные пропорции между экземплярами с положительной (был пожар) и отрицательной (пожара не было) меткой. В противном случае с моделью случится переобучение (overfitting) в сторону превалирующего класса. Из четырёх протестированных алгоритмов лучшие показатели продемонстрировали ансамблевые алгоритмы: **RF** (precision = 0.903), bagging of decision trees (precision = 0.754) и boosting of decision trees (recall = 0.658).

Подтверждают эффективность Деревьев решений и Shang, X. И Chisholm, L.A.: авторам удалось получить карту распространения пород эвкалипта с точностью более 94 % [91].

К особенностям зарубежных исследований можно отнести использование лидарных данных [25, 90]. Для зарубежных систем инвентаризации лесов применение лазерного сканирования не является исключительным, и, по словам Annika Kangas, прочно вошло в лесную практику скандинавских стран с середины 2000-х гг. [92]. Здесь надо заметить, что в США и европейских странах данные воздушного лазерного сканирования (ВЛС) имеются в открытом доступе, в отличие от России, где лазерные съёмки в целом также успешно развиваются [93-96]. В российских компаниях, в частности, в компании «Геоскан», трёхмерные облака точек получают и фотограмметрическими методами, но масштаб покрытия значительно уступает лазерному сканированию.

2.5. Становление применения МО в российском лесном секторе

В ходе изучения отечественного опыта применения МО нами была выявлена малая освещённость этого направления в отечественных публикациях. Применение машинного обучения в лесном хозяйстве России делает первые шаги, находится на стадии становления.

Прежде всего, стоит отметить работу Д. М. Черниковского: ссылаясь на успех применения k-NN в других странах, автор оценил возможности метода для определения обобщенных характеристик лесов одного из районов Ленинградской области [23]. Несовершенство национальной (государственной) инвентаризации лесов, по мнению автора, состоит главным образом в слабом использовании ДДЗ. В отличие от иностранных работ авторы статьи освещают процесс формирования набора тренировочных участков для классификации средствами геоинформационных технологий. Также отмечено, что несмотря на иные чем в канадском исследовании входные параметры (в отечественном проекте – это набор спектральных каналов Landsat-8 {GREEN, RED, NIR, SWIR 2}, а в

канадском – {EOSD, несколько каналов Landsat TM, Elevation (CDEM), Climate moisture index, Soil moisture index, Inverse distance}), алгоритм подтвердил свою действенность: была успешно выполнена автоматическая классификация по ряду лесотаксационных характеристик – среднему запасу на 1 га, среднему классу бонитета, средней относительной полноте, доле площади хвойных и лиственных насаждений. После сравнения классификации с материалами лесоустройства, метод k-NN был признан перспективным и в условиях смешанных лесов Европейской России.

Описываемая выше работа опирается на обширный обзор отечественных публикаций, при этом упомянута лишь одна работа по оценке лесопатологического состояния лесов [97].

В отечественной практике часто обращаются к космическим снимкам Landsat. С одной стороны, преимуществами использования данных снимков являются их некоммерческий характер и наличие 30-летнего ряда наблюдений с многократным покрытием крупных регионов; с другой стороны, очевидным недостатком – пространственное разрешение, уступающее новейшим платным ресурсам и лимитирующее возможные для оценки характеристики.

А. Г. Терехов, Н. Г. Макаренко, И. Т. Пак по снимкам QuickBird выполнили распознавание редколесья, чтобы потом создавать карты полноты леса [98]. Авторы видят в разработанном ими алгоритме автоматического распознавания редколесья и построения карт полноты соснового леса полезный инструмент для составления объективной оценки состояния древостоя и его изменениях. Проблема, решенная авторами, заключалась в слишком высоком разрешении снимков по сравнению с принятыми в лесоустройстве (итоговый результат они аппроксимировали на сетку с периодом 30 м), а полученные карты полноты, по мнению авторов, могут быть использованы для создания эталонов для решения подобных задач, но на основе данных, например, Landsat.

Крылов А. пишет, что по материалам Landsat возможна и идентификация ветровалов, воздействующих на устойчивость лесов [99]. Спектральная яркость ветровалов и участков неповрежденного леса различна. Это, а также форма ветровалов на снимке – полосы, вытянутые в одном направлении, дает возможность их дешифрировать. При этом чтобы исключить вырубку и гари использовали несколько разновременных снимков. В работе применялись алгоритмы SVM и Дерево решений. Первый предполагает отнесение тестового образца в один из двух возможных классов (ветровал/не ветровал), тем самым ограничивая подход только на выявление одного искомого нарушения. Карту ветровалов и оценку их площади удалось получить спустя месяц, тогда как визуальный поиск по территории в несколько миллионов десятков гектар потребовал от сотрудников

ФБУ «Рослесзащита» существенно больших усилий и времени. Соответственно, авторы в целом положительно отзываются о применении снимков Landsat для цели нахождения ветровалов, но также отмечают, что для более точного выявления участков требуются эталоны, полученные со снимков более высокого пространственного разрешения или данные наземных обследований.

Выявление и анализ опыта применения МО в лесном секторе показывает, что наиболее применимыми в настоящее время моделями являются k-NN, SVM (машины опорных векторов), Дерево решений, главным образом, ансамбль Деревьев решений – Random Forest, а также искусственные нейронные сети и глубокое обучение (ANN и DL).

Глава 3. Изученность, использование, мониторинг бореальных лесов Северо-Запада России

В работе в качестве объекта приложения применения данных ДЗЗ с машинным обучением исследуются бореальные леса Северо-Запада России. Они располагаются частично в Республике Карелия, Архангельской и Мурманской областях. Данная территория обладает высокой репрезентативностью, а рассматриваемые типы лесных экосистем распространены также в Швеции, Финляндии (т.н. Фенноскандии), Канаде и на севере США [100, 101, 92]. Эти страны являются лидерами по показателям ведения лесной отрасли, соответственно, материалы и достижения их мониторинга состояния лесов можно брать во внимание. Схожесть фенологических циклов деревьев, сезонных изменений географической среды и климатических условий может быть преимуществом при тестировании алгоритмов на данных о лесных массивах.

Выбор данной территории мотивирован следующими причинами:

- 1) большое число районов лесозаготовок (как интенсивной, так и перспективной);
- 2) наличие данных на сходные по природным свойствам территории государств, где публикуется больше открытых данных;
- 3) существенная экономическая и социальная значимость леса в регионе;
- 4) высокая изученность территории и пород деревьев (для более прозрачного конструирования признаков) за счет как отечественных, так и зарубежных исследований.

Именно природно-географическое соответствие исследуемых российских лесных экосистем аналогичным в других регионах мира служит основанием возможности опоры на зарубежные исследования при конструировании моделей МО. Таким образом формируется географический (пространственный) аспект задачи.

В главе мы во многом опирались на Лесные Планы Карелии и Архангельской области, а также на работы Крылова А. М. и Hanewinkel M [77, 78, 99, 124].

3.1. Общая характеристика лесов

Северо-Запад России – географический регион – входит в состав Северо-Западного федерального округа, граничащего с Норвегией, Финляндией, Эстонией, Польшей, Белоруссией, Латвией и Литвой. Площадь округа превышает 1,6 млн. кв. км (около 10 % территории РФ); в округе проживает около 10 % населения РФ.

Регион располагается в пределах Восточно-Европейской равнины – это низменность, всхолмленная моренными грядами – следами деятельности ледника.

Выгодное географическое расположение обусловлено выходам к Северному Ледовитому океану и Балтийскому, Белому, Баренцеву морям.

Природа региона очень разнообразна, что определяет его значительную природно-ресурсную базу [102]. На северо-востоке и шельфе доминирует нефтедобыча, тогда как главными ресурсом западной части остаются лес и минеральное сырье.

Природно-ресурсная база, соседство с развитыми странами Европейского Союза и выход к двум океанам обуславливают направления специализации округа: промышленное производство, транспортные услуги и инфраструктура, химическая промышленность, **лесопрмышленность**. Округ обладает значительным потенциалом, связанным с возможностями освоения арктической зоны.

Прогноз долгосрочного социально-экономического развития РФ предполагает, что существенное влияние на показатели развития округа будут оказывать дальнейшее развитие в регионе нефтедобывающей промышленности и сопряженных производств, но также учитывает **сохранение значения округа в развитии лесопромышленного комплекса**. В перспективе предполагается формирование в округе (среди прочих) лесопромышленного кластера - **в Республике Карелия, Архангельской области и Республике Коми** [103].

Северо-Запад России покрыт в основном таежными или бореальными лесами, для которых характерно преобладание хвойных пород (рис. 3.1). Интересы исследования обращены к таежным и северо-таежным лесным районам (в соответствии с лесохозяйственным районированием равнинных лесов Европейской части России) [104]. Леса являются основным биотическим компонентом ландшафтов. Основные лесообразующие породы: сосна обыкновенная, ель европейская и береза повислая [105]. Строение и динамика лесных сообществ регулировалась пирогенным фактором на протяжении, по крайней мере, последних тысячелетий.

Карелия и Архангельская область.

Климат региона - Республики Карелии (РК) и Архангельской области (АО) - складывается под влиянием воздушных масс, формирующихся над Атлантикой и Арктикой, характеризуется резкой изменчивостью метеорологических показателей (частая смена воздушных масс при интенсивной циклонической деятельности). Почвы преимущественно подзолистые и болотные, в АО глеево-подзолистые. В последнее десятилетие как в РК, так и в АО метеостанции отмечают увеличение средней годовой температуры воздуха. Высокая лесистость территорий определяет преимущества региона по развитию лесной отрасли (табл. 3.1).

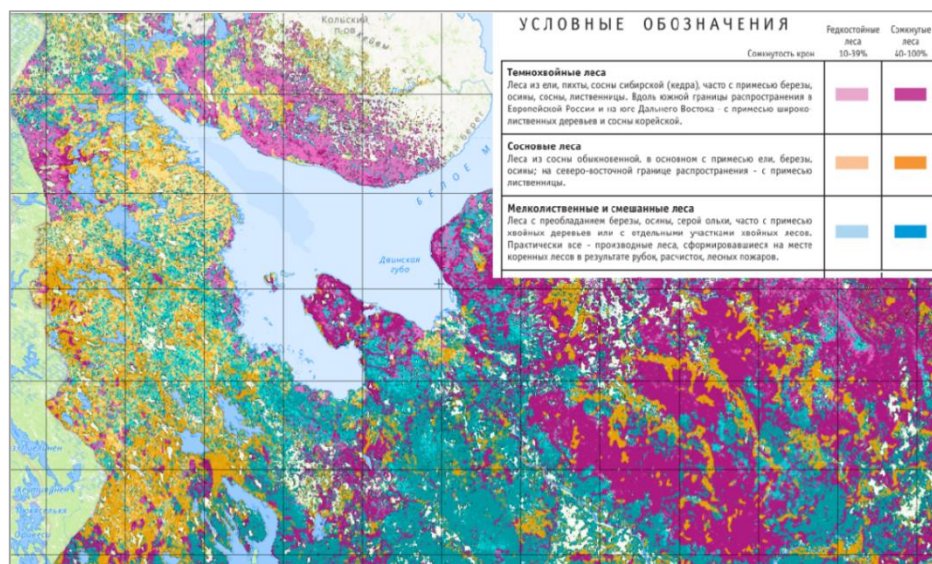


Рис. 3.1. Карта преобладающих групп пород деревьев Северо-Запада России (фрагмент интерактивной карты <https://forest.kosmosnimki.ru/>). Доминирующими породами Северо-Запада являются сосна и ель.

Табл. 3.1. Количественные показатели по лесам РК и АО [77, 78]

	Доля ЛФ, %/млн.га	Лесничества	Лесистость, %	Доля ЛПК, %	Эксплуатационные леса*, %	Ср. прирост, м ³ /га (таежный)
РК	83/14,9	23	53	30	66,4	1,50
АО	71/29,3	34	73	46	67,5	1,81

Структурная диаграмма на рис. 3.2. описывает породный состав леса Республики Карелия и Архангельской области.

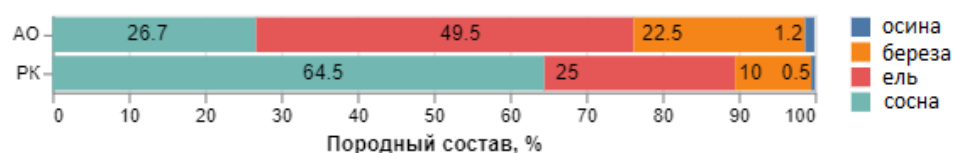


Рис. 3.2. Диаграмма породного состава ЛФ Карелии и Архангельской области (составлена по данным Лесных планов регионов).

Совокупность климатических и почвенно-гидрологических условий изменяется в неблагоприятную для роста и развития продуктивных лесов сторону в направлении с юга на север.

3.2. Экономическая значимость лесов Северо-Запада России

В настоящее время на пространственную структуру лесного покрова европейской тайги существенное, а иногда и определяющее влияние накладывает лесохозяйственная и лесозаготовительная деятельность.

На Северо-Западе РФ сконцентрированы экономически значимые леса – доля эксплуатационных лесов более 60%, при показателе по России – 51 % [106]. Лесные участки, находящиеся в федеральной собственности в составе земель лесного фонда (ЛФ), передаются в аренду лесопользователям (рис.3.3). Строительство и эксплуатация объектов для переработки древесины и иных лесных ресурсов могут осуществляться на землях ЛФ без перевода их в земли иных категорий. Согласно данным Государственного Лесного реестра на 2019 г. в Карелии арендовано около 9,8 млн. га (67 % ЛФ Республики) 65 организациями [107].

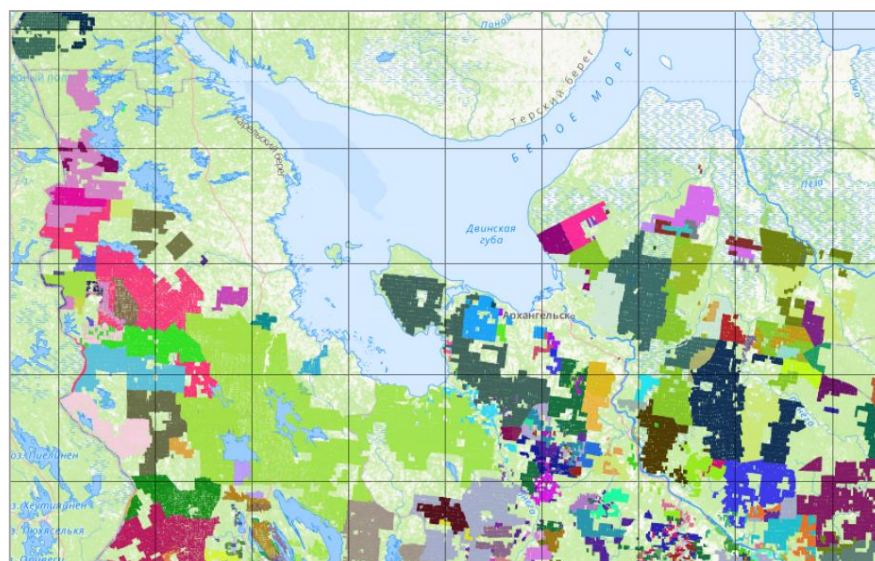


Рис. 3.3. Фрагмент интерактивной карты «Ценные леса России», слой «Арендаторы ЛФ для заготовки древесины» [108]. Распространенность предприятий – потребителей древесины на Северо-Западе РФ можно посмотреть на Демо-версии карты [109].

Лесопользователи имеют право осуществлять перечень разрешенных на конкретной категории лесов действий, обеспечивая при этом многоцелевое и рациональное использование лесных продуктов для покрытия потребностей местной экономики и населения в древесине и других ресурсах [58]. Для обеспечения соблюдения надлежащего лесопользования требуется применение передовых технологий.

В состав лесной промышленности входят отрасли, занимающиеся добычей, переработкой древесины, а также производством готовой продукции. Описание видов деятельности предприятий приводится в табл. 3.2.

Табл. 3.2. Отрасли ЛПК и деятельность предприятий данной отрасли [66, 110].

Лесозаготовка	Деревообработка	Целлюлозно-бумажная промышленность	Лесохимическая промышленность
Вырубка и первичная обработка древесины.	Обработка древесины, изготовление стройматериалов, мебели и т.д.	Высокотехнологичная отрасль. Основной ресурс - хвойные породы, (распространенность хвойных на СЗ России (рис. 3.1.) обуславливает распространенность предприятий на территории региона [110]. Наносит огромный вред экологии.	Потребляет отходы производства древесины (опилки, стружку), располагается ближе к районам заготовки и переработки леса. Продукты – целлофан, линолеум, искусственное волокно [66].

Лесопромышленный комплекс Карелии является ведущим в экономике республики и характеризуется развитием таких видов деятельности, как лесозаготовки, деревообработка и целлюлозно-бумажное производство [77]. Ведущее место по объему производства в АО занимают целлюлозно-бумажное производство и деревообработка (около 46 %) [78].

Предприятия ЛПК относятся к фондоёмким и материалоёмким типам бизнеса [111]: нормальное функционирование заготовки и переработки древесины требует специальной техники, транспортных средств и пр., и конечно, самой древесины.

В регионе показатели использования древесины высоки, но темпы вырубки значительно превышают темпы воспроизводства ресурсов. Поэтому вопрос о рационализации лесохозяйственной деятельности стоит остро. По показателям расходов на лесовосстановление, отнесенным к площади лесных земель конкретного округа (табл. 3.3), СЗФО уступает только Дальневосточный ФО (рис. 3.4). В то же время на охрану, защиту и воспроизводство леса на одном гектаре в СЗФО тратится средств на порядок больше.

Табл. 3.3. Затраты на лес (составлена на основе данных Рослесхоза [64, 112, 113]).

ФО	Площадь лесных земель, тыс.га	Затраты на лесовосстановление, руб/га	Затраты на охрану и лесовоспроизводство, руб/га
Центральный	23392	22.27	511.23
Северо-Западный	93209	0.76	186.82
Южный	3049	46.69	695.69

Северо-Кавказский	1731	22.78	351.52
Приволжский	39023	7.12	324.41
Уральский	72025	1.97	93.69
Сибирский	243848	2.08	87.54
Дальневосточный	417859	0.65	29.72
Карелия	9784	0.73	281.89
Архангельская обл.	22769	0.90	213.53

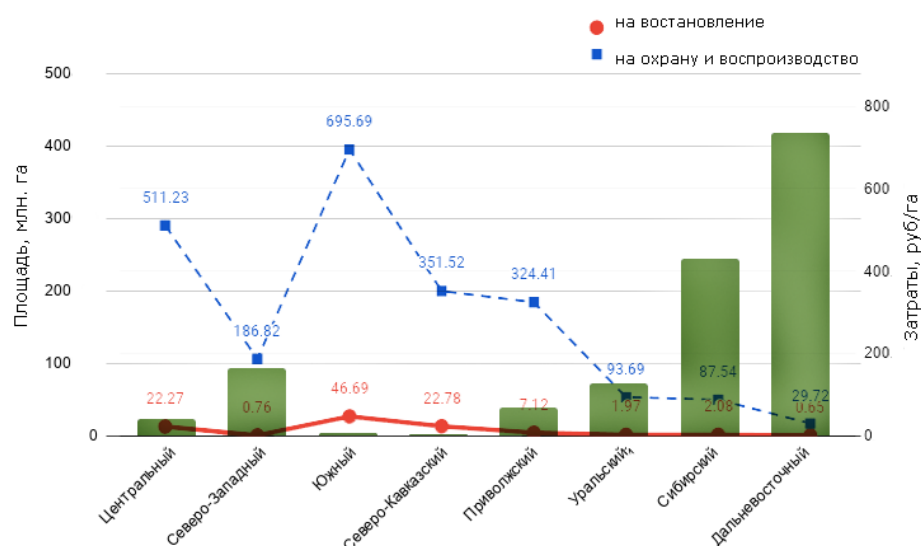


Рис. 3.4. Затраты на лесовосстановление и на охрану и воспроизводство лесов (по правой оси) на фоне размеров лесных площадей регионов РФ (по левой оси). 2018 г.

За этот же период объем заготовленной древесины в СЗФО превысил 60 тыс. м³ – четвертую часть лесозаготовок в целом по стране [69]. Приведенная статистика подчеркивает место региона в лесном комплексе страны и социально-экономическое отрасли в регионе.

Однако по данным, размещенным на сайте Федерального Агентства Лесного Хозяйства, результаты плановых и внеплановых проверок, проведенных Рослесхозом на территории Северо-Запада, помимо общей халатности в ведении документации, свидетельствуют также о недостаточном объеме санитарно-оздоровительных мероприятий (Архангельская обл. и Республика Коми). Отмечается также, что в Архангельской области не выполняется расчетная кратность авиационного патрулирования, что объясняет недостаток материалов для мониторинга ситуации и принятия решений [114].

В условиях развития рыночных отношений постоянно возрастает актуальность проблемы рационального лесопользования.

Для Северо-Запада России, где «через стенку» передовые в плане ведения лесного хозяйства Финляндия и Швеция, недостатки лесопользования видятся весьма отчетливо. При близких средних эксплуатационных запасах величина ВВП лесного сектора на гектар эксплуатационной площади в год в России намного ниже, чем в Финляндии и Швеции (рис.3.5). В разработанной Санкт-Петербургским НИИ лесного хозяйства (ФБУ «СПбНИИЛХ») Концепции [115] говорится о происходящей долговременной смене хвойных пород на лиственные, в результате чего большая часть лесоперерабатывающих компаний испытывают дефицит сырья.

Показатель	Россия	Швеция	Финляндия	США	Канада
Лесная площадь, млн га	809	28	22	304	310
Эксплуатационные леса, млн га	494	25	20	231	273
Запас на корню, млн м³	81 523	3358	2189	47 088	32 983
Средний запас на гектар, м³	101	119	99	155	106
Годовой ВВП лесного сектора, млрд \$	6,8	12,8	10,3	108,4	32,0
То же в расчете на гектар эксплуатационной площади, \$/га	14	508	512	469	117
Вклад лесного сектора в общий ВВП, %	0,8	3,8	5,7	0,8	2,7

Рис. 3.5. Эффективность лесного сектора в некоторых лесных странах [116].

Шварц Е.А. и Шматков Н.М. дают общий вывод по публикациям, посвященным анализу недостатков используемых методов исчисления расчетной лесосеки и предложениям по их совершенствованию: расчетная лесосека, определяемая по используемым методикам, не позволяет обеспечить непрерывное и устойчивое лесопользование, в частности потому, что не учитывает экономическую доступность лесных ресурсов и основана на недостоверных, как правило, устаревших, лесоустроительных данных [116]. Именно эта ситуация подкрепляет в нас уверенность необходимости наших исследований.

Альтернативой экстенсивной модели является интенсивная модель лесного хозяйства, вопрос о переходе к которой обсуждался давно и наконец нашел отражение в лесной политике: в 2019 г. в Карелии начали внедрение модели интенсивного использования и воспроизводства лесов [117]. Неотъемлемые составляющие модели — **сбор и анализ информации о насаждениях**, **рыночная оценка ресурсов** и прогнозирование рыночного потенциала различных сортиментов, эффективное лесовосстановление и активное **использование рубок ухода** с формированием древостоев желательной породной, возрастной и товарной структуры [118]. Переход поддержали крупные «игроки» отрасли, например, ОАО «Сегежа». В ходе внедрения интенсивного лесного хозяйства (ИЛХ) компании должны проводить оценку всей арендной базы, выявляя структуру и состав насаждений.

Этому переходу предшествовало накопления опыта имитации естественной динамики лесов. В тесном сотрудничестве со Шведским лесным агентством осуществлялся «Псковский модельный лес» – проект Всемирного фонда дикой природы (WWF). К созданию концепции «модельного леса», примера устойчивого лесопользования, привело большое количество острых дискуссий вокруг нерационального использования природных ресурсов лесной промышленностью в 1980-х годах в Канаде [119]. На тот момент (2009) заготовки древесины в Швеции уступали российским цифрам всего в два раза (примерно 100 млн м³ против 200 млн м³), тогда как площадь лесов в РФ более чем в 30 раз больше. Скандинавы признали важность российского лесного сектора для Швеции: есть чему поучиться у разработчиков проекта, особенно в части сохранения биоразнообразия, ведь в Швеции осталось мало девственных лесов и приходится работать над его восстановлением [120]. Это не единственный пример «скандинавского влияния» в отечественном лесном секторе.

По заключению группы ученых из Швеции и других стран Балтийского региона [56] на северо-западе России есть возможности для интенсификации лесопользования. В статье приводится графика соотношения потребления древесины и биоразнообразия для пяти тестовых территорий, и видно, что районы исследований на северо-западе России продемонстрировали самые высокие уровни биоразнообразия, тогда как экономические показатели были низкими (рис. 3.6).

О сравнении биоразнообразия в пользу России говорят и данные исследования, цитируемого Шматковым Н.М. в статье «"Скандинавская" Модель Интенсивного Лесного Хозяйства» [121]. Автор делает вывод, что восстановление естественной флоры и фауны очень трудо- и время-затратное дело, несмотря на активность предпринимаемых скандинавами мер.

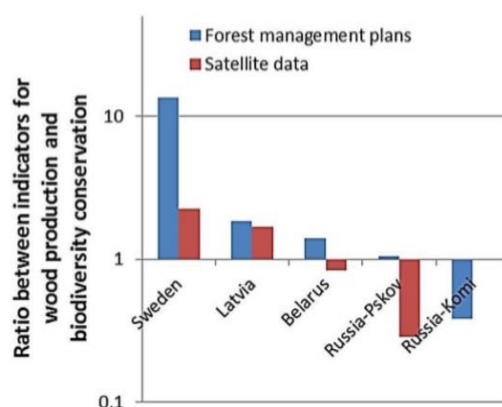


Рис. 3.6. Соотношение интенсивности ЛПК с биоразнообразием на тестовых территориях. Пространственные данные получены исследователями из лесных планов и с открытых систем ДДЗ [56].

Стоит отметить и успехи лесного сектора Северо-Запада России: по данным на начало 2020 г. Россия обогнала Канаду, много лет лидировавшую по площади лесных территорий, на которых ведется ответственное лесоправление по схеме FSC (сертификат Лесного попечительского совета или FSC (Forest Stewardship Council) – эффективный инструмент поддержки экспорта отечественной лесопроductии, на продукцию с логотипом FSC есть устойчивый спрос на экологически чувствительных рынках Европы, Северной Америки, Азии): в этом регионами-лидерами стали Архангельская область, Республика Коми, Вологодская область, Республика Карелия и Ленинградская область [122].

По мнению Шматкова Н.М. [121], директора представительства FSC России, интенсификация лесного хозяйства в староосвоенных регионах позволит преодолеть наступающий кризис обеспечения лесной промышленности сырьем и сохранить ценные массивы малонарушенных лесов.

Как было отмечено, реализация механизма интенсивной модели предусматривает и экономическую оценку. С переходом к рыночной экономике в РФ появился большой пласт различных видов менеджмента, которые включены в системы управления, в частности лесопользования. Риск-менеджмент в лесном хозяйстве должен учитывать и оценивать специфические отраслевые риски: лесные пожары, размножение вредителей, болезни леса, неблагоприятные погодные условия (ветровал, засуха и т.д.), незаконные рубки леса и другие воздействия (загрязнения промышленными выбросами и т.д.) [123]. Перечисленное является основными причинами крупных нарушений лесных экосистем в бореальной зоне. Для моделирования экономических и экологических рисков обязательны учет количественных показателей. В лесном хозяйстве не все вопросы управления рисками являются достаточно изученными и разработанными, особенно с учетом региональных особенностей.

Заболевания.

Массовые вспышки насекомых могут привести к серьезным утечкам углерода и могут вызывать петли обратной связи с опасностью дальнейшего ухудшения ситуации в целом [124].

Согласно Schelhaas et al. [125] за период 1950–2000 гг. 2,9 млн. м³ поврежденной древесины в год можно приписать деятельности короедов. Другие насекомые, которые исторически наносили большой ущерб, в настоящее время не оказывают такого же воздействия, как короеды. Лесные насаждения оказываются более уязвимы к вспышке короеда после пожаров и наоборот [124], также выявлена четкая взаимосвязь между дефолиацией насекомых и пожарами на ландшафтном уровне. Модели взаимодействия

между возмущающими агентами, такими как лесные и еловые жуки [124], трудно количественно оценить по величине и по времени и, следовательно, включить в модели риска [124].

Д.Ф. Налдеев изучал массовые вспышки размножения жука-короеда в НП «Водлозерский» РК 2000-х гг. [126]. Крылов А.М. проводил работы по выявлению очагов короедов с использованием снимков Landsat на территории Московской области [127].

Ветровалы.

В регионе резкие смены ветра, и сильные ураганные ветра наносят урон лесам, т.к., как было сказано ранее, территория находится в зоне активной циклонической деятельности. Многие исследования подтверждают, что особенно после штормовых явлений с ветровалами происходят вспышки размножения насекомых в результате предыдущих абиотических нарушений [128].

В статье А.М. Крылова и А.А. Соболева [127] приведена выявленная типичная картина таких повреждений лесов ветровалами европейской части России – вытянутые многокилометровые полосы (рис. 3.7).

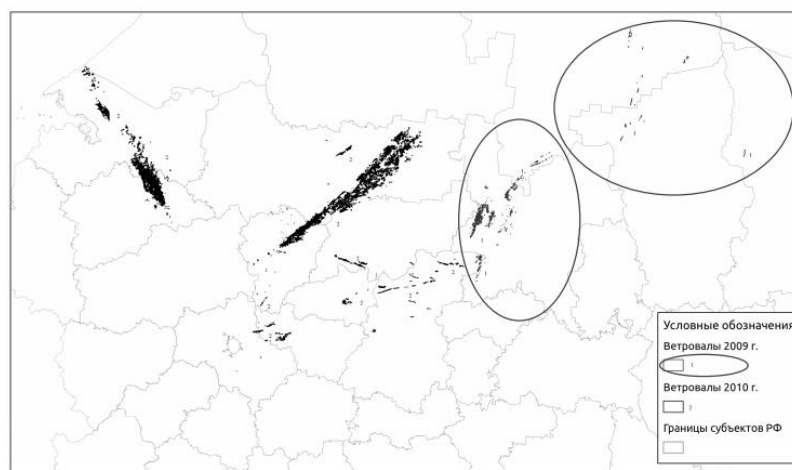


Рис. 3.7. Структурная картина ветровалов на мелкомасштабной карте Европейской России [127].

В рамках данной работы рассматриваемые лесные массивы располагаются, в том числе в пределах особо охраняемых природных территорий (ООПТ). Они в настоящее время не эксплуатируются экономически, однако в условиях кризиса лесобеспечения могут оказаться под угрозой. Следует принять во внимание, что ведение лесоучетных работ и виды лесоучетных материалов защитных лесов и эксплуатационных лесов не имеют принципиальных отличий.

3.3. Действующие системы учета лесов Северо-Запада России

На территорию Северо-Запада России распространяется действие лесного законодательства РФ, методы таксации, оценки, проведения ГИЛ и мониторинга. По мнению А. Марковского и А. Родионова действующее лесное законодательство России не содержит в себе ограничений или препятствий, делающих невозможным ведение интенсивного лесного хозяйства на арендованных участках лесного фонда [129].

Учет лесов Северо-Запада России проводится согласно статье 90 ЛК РФ «Государственная инвентаризация лесов» [58]. В отношении лесничеств и лесопарков ведется Государственный лесной реестр (ГЛР).

На землях лесных фондов Республики Карелия и Архангельской области действуют, соответственно, 23 и 34 лесничества. Как основные территориальные единицы управления в области защиты и воспроизводства лесов лесничества образованы и на землях ООПТ. Лесохозяйственные регламенты лесничеств на ООПТ доступны по [130].

Различные виды работ по учету и оценке лесов делегируются региональным подразделениям Рослесхоза, лесничествам. Материалы лесоустройства создаются в электронном виде в формате совмещенной векторной и атрибутивной базы данных. Собранная информация передается в ГЛР, также используется для расчетно-аналитического планирования с привлечением действующих методик лесотаксационных расчетов, экономической оценки лесов, оценки лесосырьевых ресурсов, расчета бюджета углерода.

Информацию о ныне применяемых методиках расчетов и оценок ключевых показателей можно найти в лесных планах субъектов. Так, методика оценки углеродного баланса полагается на расчет углерода надземной фитомассы древостоя и осуществляется через приводимые в материалах ГЛР объемные запасы древесины насаждений (также косвенная характеристика) и ряд конверсионных коэффициентов [Приложение 6 из 77]. Поскольку бюджет углерода суммируется по всем пулам (живой древостой, древесный отпад, подстилка, почва), то вовлекается и количественная информация о масштабе «отпавшего» леса.

При заполнении ГЛР одним из первичных документов являются таксационные описания (табл.3.8). В Республике Карелия атрибутивная по выделу база данных, не привязанная к электронным картам, ведется в Автоматизированной системе управления лесными ресурсами (АСУЛР). Она позволяет вносить изменения на выделном уровне. Изменения в таксационные описания и лесоустроительные планшеты на бумажных носителях вносятся специалистами лесничеств [77]. В Архангельской области формы ГЛР предоставляют на бумажных носителях и в электронном виде [78].

Табл.3.4. Фрагмент таблицы «Изменение таксационных характеристик лесных насаждений по лесничествам и их анализ за период действия предыдущего Лесного плана Республики Карелия» [77].

Наим-е лесничества, лесопарка*	Покрытая лесом площадь, га.	Ср. возраст, лет	Ср. класс бонитета	Ср.относительная полнота	Средний запас насаждений на 1 га м ³		Ср прирост по запасу на 1 га покрытых лесной растительностью земель, м ³	Состав насаждений
					земель, покрытых лесной растительностью	спелых и перестойных		
1	2	3	4	5	6	7	8	9
Беломорское 01.01.2009	647752	63	4,4	0,65	78	122	1,5	5С2Е3Б
Беломорское 01.01.2018	650179	68	4,3	0,65	88	126	1,6	6С2Е2Б
Изменения 2009 -2017 гг.	2427	5	0,1	-	10	4	0,1	+1С; -1Б
Калевальское 01.01.2009	775685	87	4,6	0,56	81	135	1,0	7С2Е1Б
Калевальское 01.01.2018	779390	85	4,6	0,56	78	134	1,0	7С2Е1Б

Применение данных ДЗЗ и автоматическое дешифрирование в учете лесов в настоящее время не может быть названо основным методом получения информации о состоянии лесных земель. Дешифровочный способ таксации разрешен к использованию в отношении эксплуатационных и защитных лесов в случае, если назначения мероприятий в лесных насаждениях не требуется. Дистанционные наблюдения в целях мониторинга здоровья лесов и форму их результатов регламентирует Порядок осуществления государственного лесопатологического мониторинга [131]. Данные нормативный акт допускает привлечение космических снимков с разрешением 1-30 м (в исключительных случаях до 100 м) и применение автоматических методик дешифрирования (с привлечением результатов наземных наблюдений на тестовых участках). Для снимков разрешения 10-40 м минимальная фиксируемая площадь нарушения составляет 10 га, а происхождение скопления мертвого леса не определяется. Если используются снимки разрешением 3-10 м, то изменения в санитарном состоянии должны фиксироваться на лесных участках площадью от 1 га. Распознаваемыми структурными деталями являются выделы (повыдельная фиксация).

В труднодоступных для проведения наземных работ участках и при слабой лесопатологической угрозе методы ДЗ используются для обобщенной оценки площади погибших и поврежденных лесных насаждений. Ликвидация своевременно выявленного сухостоя, к слову, снижает потенциальную горимость леса за счет уменьшения количества горючих материалов. Результатами дистанционных наблюдений за санитарным состоянием лесов должны являться специальные карты (масштаб не мельче 1:50 000, растровый формат), на которые должны быть нанесены контуры лесных насаждений с выявленными изменениями в лесопатологическом состоянии.

Социально-экономические особенности Северо-Запада России определяют высокая степень обеспеченности природными ресурсами и территориальная близость к зарубежным (европейским) лесным рынкам. Исследуемый регион относится к лесным районам, где исторически активно функционирует лесопромышленный комплекс (ЛПК). В целях повышения эффективности деятельности ЛПК и одновременного сохранения главных экологических функций лесов в регионах начали перенимать модель интенсивного лесного хозяйства, успешно практикуемую скандинавскими странами. Однако недоступность актуальной и точной информации о лесных территориях служит препятствием для качественного планирования и эффективного управления ресурсами [132]. Для мониторинга состояния лесных насаждений законодательством допускается использовать данных ДЗЗ и применять автоматическое дешифрирование с привлечением полевых данных с тестовых территорий. Мы видим здесь возможность для опробования алгоритмов машинного обучения как путь получения пространственной информации требуемой точности и подходящей для составления специальных карт.

Глава 4. Создание модели классификации бореальных лесов с деградациями

Проведенный анализ и изучение опыта работ Крылова А.М., Черниховского Д.М., Abdi A. M. по интеграции машинного обучения в отношении лесных экосистем, позволяет приступить к разработке и апробации классификации лесного покрова по данным дистанционного зондирования с применением машинного обучения. Объектом исследования является нарушенный лесной покров в бореальных лесах Карелии на Северо-Западе России, предметом исследования – применение алгоритмов машинного обучения и спутниковых данных для обнаружения и оценки деградировавших древостоев в среде ГИС и в среде разработки Python.

4.1. Цели и задачи исследования

Лесное законодательство РФ выделяет следующие подкатегории лесных земель с нарушенным древостоем (см. гл. 2.): гари (полные и те, на которых возможно естественное возобновление), вырубки (древостой сведен полностью), погибшие (от стихии, фитовредителей). Выделение последней категории – задача, решаемая в ходе лесопатологического мониторинга (ЛПМ), который в силу обширности и трудноступности условий северных регионов целесообразно проводить с применением данных ДЗЗ. Комбинирование дистанционных методов и умной обработки данных следует рассматривать как способ повышения качества обследования лесов и информирования государственных органов, крупных арендодателей и научных организаций о состоянии лесов. Полученная пространственная информация ложится в основу оценки планируемых затрат на проведение лесовосстановления; оценки обеспеченности сырьем промышленности, перерабатывающей лесные ресурсы; оценки экологического потенциала, потенциала средообразующих, водоохраных, защитных, санитарно-гигиенических, оздоровительных и иных полезных функций лесов.

Исходя из этого цель практической части настоящего исследования состоит в применении алгоритмов машинного обучения к спутниковым данным для выявления нарушений в лесном покрове, а также в оценке работоспособности полученных моделей. Географически исследование фокусируется на бореальной лесной зоне Северо-Запада России. Для создания модели МО и проведения классификации лесов, в которых присутствуют нарушения, требуется решить следующие задачи:

- сформировать набор исходных пространственных данных по лесам, в том числе сведения о точно известных местах с деградациями древостоя для создания обучающей выборки (эталонов);

- с учетом требований выбранных алгоритмов к входным данным подготовить обучающие выборки и перевести пространственные данные в машиночитаемый формат средствами геоинформационных систем;
- построить модель МО, провести обучение и валидацию, с помощью обученной модели провести автоматизированное дешифрирование - классификацию повреждений лесного покрова и оценить точность работы модели;
- визуализировать результаты автоматизированного дешифрирования и сопоставить с материалами лесоустройства, выполненного преимущественно полевыми методами;
- сделать вывод о применимости исследованных моделей и использованного набора данных для классификации повреждений в лесах и об их действенности для оценки лесных ресурсов.

4.2. Выбор алгоритмов

Для обработки ДЗЗ методами машинного обучения нами были выбраны два алгоритма классификации: SAM и Random Forest. Первый алгоритм реализован в плагине QGIS, что освобождает пользователя от необходимости конфигурировать не только алгоритм, но и его параметры. Задача классификации сводится к ручному набору обучающей выборки. Это полуавтоматическая классификация. При помощи данного плагина Карпачев А.П. осуществлял выявление повреждений лесов национального парка (НП) «Орловское полесье» [133]. Второй алгоритм Random Forest был выбран на основе рассмотрения главным образом зарубежных публикаций. Обучающая выборка на основе данных лесопатологического обследования (ЛПО) для создания модели Random Forest была подготовлена в QGIS, а сама модель написана на языке Python. Много внимания классификации лесопатологического состояния лесов можно найти в работах Крылова А.М. [99], который использовал другие алгоритмы и иную среду разработки. Тем не менее, мы во многом опирались на его опыт создания эталонов.

4.3. Формирование набора исходных пространственных данных

В реализации машинного обучения ключевую роль играют данные, вовлекаемые в обработку. Поэтому проектирование и создание базы данных нарушений составляет значительный и значимый объём работ.

Для проведения классификации лесного покрова с нарушениями требуются данные о поврежденных участках (пожарами, болезнями, ветрами, ураганами). Возник вопрос о наличии, качестве и доступности таких данных на участки эксплуатируемых лесов – они

находятся в пользовании государственных структур и иных организаций, вовлеченных в лесохозяйственную деятельность.

В результате поиска путей получения фактических данных о нарушениях в лесных насаждениях с учетом описанных требований, был выбран вариант использования официальных открытых данных с портала Министерства природных ресурсов [134]. На сайте министерства регулярно пополняется раздел с документацией ЛПО особо охраняемых природных территорий (ООПТ). В процессе ЛПО производятся определение причин повреждений (или гибели) лесных насаждений, а также определение местоположения и границ поврежденных лесных участков. Именно информация о местоположении и типе повреждения нужна для первоначальной установки эталонных участков.

Стоит отметить, что в настоящем виде информация о ЛПО стала доступна только в 2017 г. с выходом приказа Минприроды № 480. Участники встречи, состоявшейся в феврале 2017 г. – представители Гринпис, WWF, руководства Минприроды и Рослесхоза - пришли к выводу о необходимости изменений, предусматривающих опубликование актов ЛПО как минимум вместе с абрисами лесосек, с исключением информации, доступ к которой ограничен действующим законодательством [135].

Минприроды публикует акты обследований, проводимых, в том числе, на части территории ООПТ. Данные акты являются официальными документами и вся содержащаяся в них информация написана на языке лесоустройства. На основании -актов обследований и подсчета лесотаксационных показателей назначаются требуемые лесовосстановительные мероприятия.

В рамках сбора материалов настоящего исследования были отобраны акты, составленные по результатам обследования лесных выделов на территориях НП «Паанаярви» (Республика Карелия), НП «Водлозерский» (Республика Карелия, Архангельская обл.) (рис. 4.1). ЛПО лесных насаждений этих территорий было проведено полевыми инструментальными методами. Документов о проведении ЛПО с использованием данных дистанционного зондирования Земли (авиационных и космических снимков) обнаружено не было.

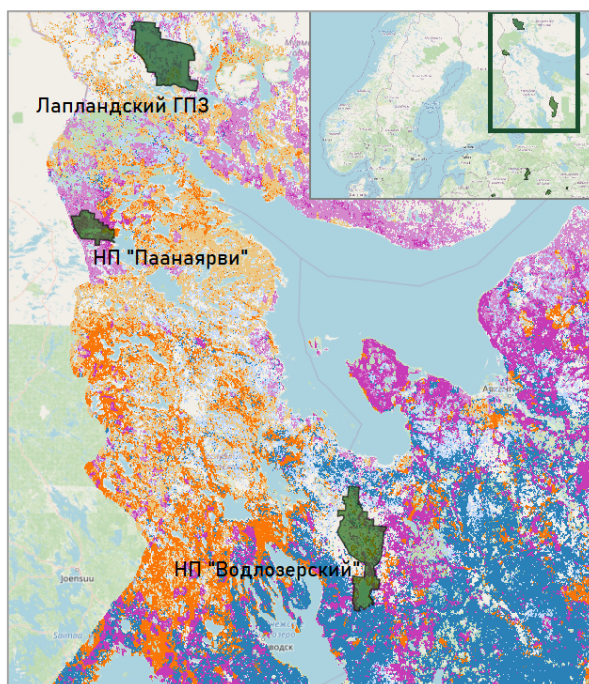


Рис. 4.1. Географическое положение территорий исследования на карте преобладающих групп древесных пород.

Несмотря на то, что земли ООПТ национальных парков изъяты из хозяйственного использования и цели при организации и ведении хозяйства на них иные, характеристики деревьев и, следовательно, намечаемые мероприятия лесовосстановления незначительно отличаются от тех лесных массивов, отнесенных к эксплуатационным. Лесоустройство на землях ООПТ производится в установленном порядке и в эти работы поэтапно внедряются ГИС-технологии [63]. Таким образом, данные, полученные с территорий лесов на охраняемых землях, **могут быть использованы в данной работе.**

На геопортале «Леса высокой природной ценности Республики Карелия» (ЛВПЦ) [136] доступна геопространственная информация по лесничествам, участковым лесничествам и кварталам лесничеств. Некоторые слои доступны для скачивания [137]: Малонарушенные лесные территории, Малонарушенные Лесные Массивы, Лесохозяйственные границы Карелии (шейп-файлы на лесничества, участковые лесничества и кварталы (без ООПТ)).

В целях исследования нас интересует слой типов покрытий (рис. 4.2). Для территории Карелии доступно растровое изображение и файл привязки .tfw (СК EPSG:3395 - WGS 84 / World Mercator - Projected).

Слой Растительный покров и антропогенные ландшафты Республики Карелия	
Актуальность	2011 год
Авторство	НП "Прозрачный Мир"
Публикация	http://hcvf.ru/ru/publications/sohranenie-tsennih-prirodnih-territoriy-severo-zapada-rossii-analiz-reprezentativnosti
Методика	на стр. 81 публикации
Легенда	1 Ель
	2 Сосна зеленомошная
	3 Сосна лишайниковая
	4 Сосна сфагновая
	5 Лиственные
	6 Смешанные
	7 Сфагновое болото
	8 Травяное болото
	9 Обводненное болото
	10 Вырубки
	11 Пожары
	12 Ветровалы
	13 Водные объекты
	14 Тундра
	15 Отсутствие растительности, вызванное антропогенным вмешательством
	16 Пашни
	17 Луга
	18 Отсутствие растительности, вызванное естественными причинами
Примечания	Обновлено НП "Прозрачный мир" по данным 2011 года

Рис. 4.2. Описание слоя «Растительный покров и антропогенные ландшафты Республики Карелия» и его легенда. [136]

Именно на эти материалы можно ориентироваться при подборе комбинации каналов мультиспектральных снимков, при наборе пикселей для обучающей выборки.

4.4. Подготовка набора данных

Создание набора пространственных данных на основе текстовых данных актов ЛПО может быть осуществлено в пять принципиальных шагов (рис. 4.3.), предпринятых: для создания свободной таблицы ключевых показателей по материалам ЛПО (1); для векторизации контуров выделов с нарушенным древостоем (2); для поиска и скачивания космических снимков по дате из акта ЛПО (3); для загрузки необходимых каналов мультиспектрального снимка в среду QGIS (4); для совмещения спутниковых данных и векторных изображений эталонных выделов в проекте QGIS в единой системе координат (5).

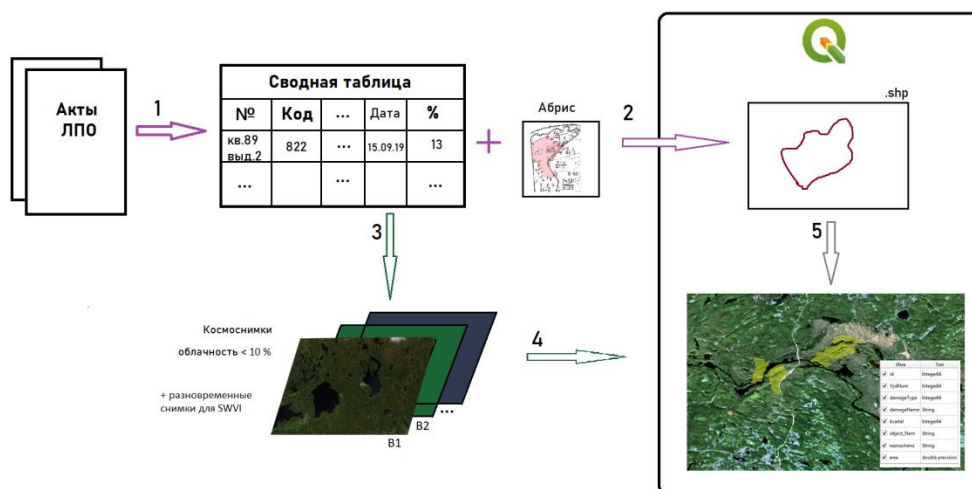


Рис. 4.3. Обобщенная схема подготовки эталонных данных.

Для создания обучающей выборки в актах ЛПО особое значение имеет раздел 2 и абрис обследованного участка в приложении. В разделе 2 указано состояние насаждений и причины повреждений с кодами, принятыми в таксационном деле. Причины повреждений представляются по признакам, которые также имеют коды. Так, воздействия шквалистых ветров имеют код 822, повреждения короедом – код 343 и др.

В ходе анализа актов ЛПО из них извлечена необходимая информация в инвентаризационную таблицу (табл. 4.1). Графа «Процент нарушенного леса» является ключевой, она будет нужна для проверки качества работы алгоритма автоматической классификации повреждений лесов (т.н. ground truth).

Табл.4.1. Характеристики эталонных территорий с зафиксированными повреждениями лесов (фрагмент).

	Объект	Место	Дата ЛПО	Площадь ЛПО/Площадь выдела, га	Тип нарушения	Процент наруш. леса, %	Причина	Порода	Примечания
1	НП <u>Паанаярви</u>	89 квартал, 1 выдел (рекреационная зона, участок 3А)	15.12.2019	66/66	Ветровал	13	Шквалистые и ураганные ветра прошлых лет		13 % старого сухостоя Назначено ВСП, естественное возобновление
		63 квартал, 49 выдел (рекреационная зона, участок 3А)	15.12.2019	47/47	Ветровал	23	Шквалистые и ураганные ветра прошлых лет		23 % старого сухостоя Назначено ВСП, естественное возобновление
		90 квартал, 6 выдел (рекреационная зона, участок 3А)	15.12.2019	116/116	Ветровал	10,5	Шквалистые и ураганные ветра прошлых лет		10,5 % старого сухостоя Назначено ВСП, естественное возобновление
		93 квартал, 15 выдел (рекреационная зона, участок 3А)	15.12.2019	18/18	Ветровал	10,5	Шквалистые и ураганные ветра прошлых лет		10,5 % старого сухостоя Назначено ВСП, естественное возобновление

Абрис обследованного участка представляет собой часть лесоустроительного планшета - первичного документа планово-картографических лесоустроительных материалов (рис. 4.4). Абрисы составляются по результатам съемки и таксации леса в масштабах 1:10 000 и 1:25 000. Они также нами будут использоваться проверки качества работы алгоритма автоматической классификации повреждений лесов.

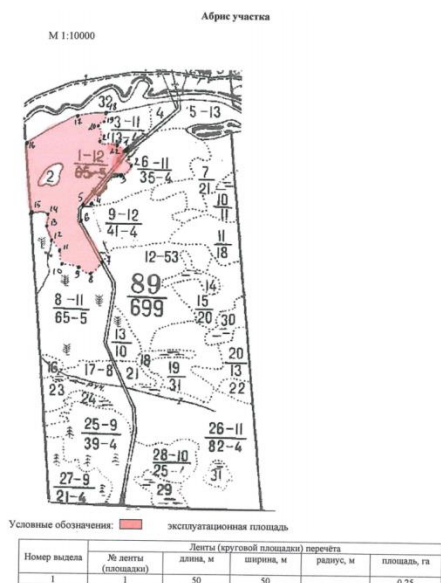


Рис. 4.4. Абрис акта ЛПО - эталонной территории в исследовании (НП «Паанаярви»).

На каждом выделе абриса приводят в виде дроби следующие показатели, идентичные показателям в таксационном описании: в числителе – номер лесного выдела и класс возраста леса, в знаменателе – площадь лесного выдела и класс бонитета (оценки леса) [138].

Сложность в распознавании конкретного участка на абрисе, избранного в качестве эталона, для его идентификации на снимке чаще всего связана с недостатками оформления математических элементов абриса – координатных линий и их подписей, также координат характерных точек выдела, что противоречит Приказу «Об утверждении порядка проведения лесопатологических обследований и формы акта лесопатологического обследования» [139]. Он устанавливает, что для фиксации углов границ поврежденных и погибших лесных участков определение географических координат характерных точек делается при помощи навигационных приборов. Однако, вместо координат угловых точек участков к абрису приводят таблицы румбов и длин линий без координат начальных точек. Для опознания и определения местонахождения эталонного выдела на снимке необходимо обращаться к материалам лесоустройства, выявлять и сопоставлять характерные формы рельефа и названия объектов абриса с другими картографическими материалами.

Лесоустроительные материалы на ООПТ сложно найти в открытом доступе. В лесохозяйственных регламентах лесничества на территории национальных парков приводятся данные о разделении на рекреационную, заповедную зоны, но не о повыведельном разделении территории. Для определения местоположения конкретных выделов с абрисов ЛПО привлекались различные средства, от лесоустроительных материалов и приказов об учреждении национальных парков до растровых схем лесоустройства (картинок, фотографий), найденных на сайтах соответствующих парков. Так например, для поиска местоположения эталонных участков на территории НП «Паанаярви» было использовано распоряжение по НП с указанием кварталов, отнесенных к заповедной, особо охраняемой и рекреационной зонам и название озера на одном из абрисов.

Создание векторного файла с участками с нарушениями проводилось методом ручной векторизации в открытом ПО **QGIS 3.10**. Векторизация границ выделов осуществлялась по заранее привязанным к топооснове растровым изображениям абрисов. Здесь надо заметить, что существуют специализированные коммерческие геоинформационные системы (например, ГИС ТороL-L), адаптированные под задачи лесоустройства, обладая лицензией на которые, ручной векторизации не потребовалось бы.

Привязка растровых изображений абрисов выполняется либо с привлечением лесоустроительных материалов, либо по характерным точкам эталонных выделов (рис. 4.5). Во втором случае детальности стандартных базовых карт было недостаточно, и в проект подгружался лист топографической карты с привязкой OZI-explorer. Найденная таким образом на подложке местность обводилась контуром. В результате каждый эталонный выдел - полигон - получил номинальное цифровое значение в зависимости от принадлежности к определенному классу повреждения и иные атрибуты, выбранные из лесотаксационной информации, приведённой в акте ЛПО (рис. 4.6).

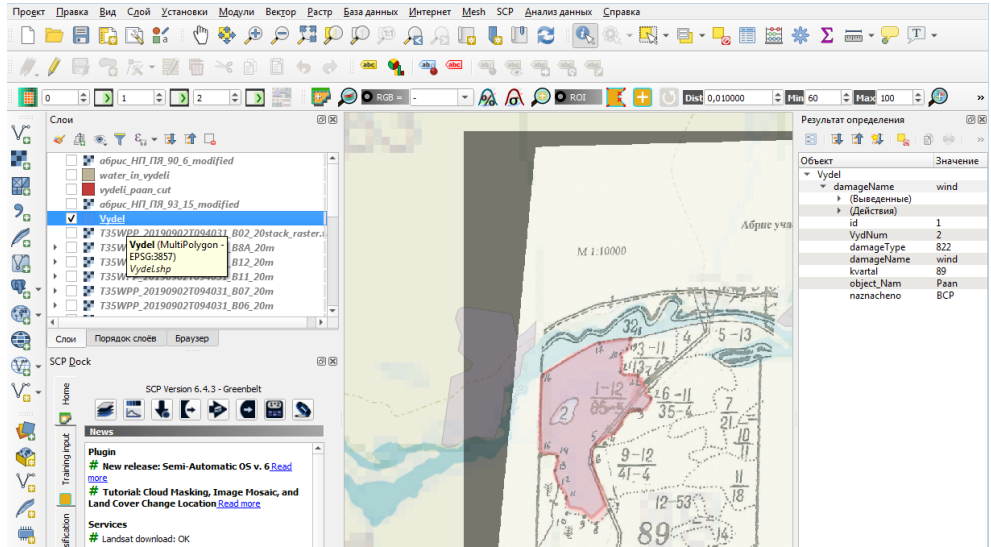


Рис. 4.5. Рабочее окно ПО QGIS - векторизация выделов для эталонной выборки.

Имя	Тип
<input checked="" type="checkbox"/> id	Integer64
<input checked="" type="checkbox"/> VydNum	Integer64
<input checked="" type="checkbox"/> damageType	Integer64
<input checked="" type="checkbox"/> damageName	String
<input checked="" type="checkbox"/> kvartal	Integer64
<input checked="" type="checkbox"/> object_Nam	String
<input checked="" type="checkbox"/> naznacheno	String
<input checked="" type="checkbox"/> area	double precision

Рис. 4.6. Атрибуты полигонального слоя выделов.

Для проверки того, что полигоны выделены верно, была определена их площадь. Она должна совпадать с площадью выдела из акта ЛПО. Для полигонального слоя эталонных выделов были также рассчитаны и добавлены в качестве атрибутов выдела статистические показатели классифицированного растра (кол-во пикселей, среднее, вариативность, СКО и т.д.).

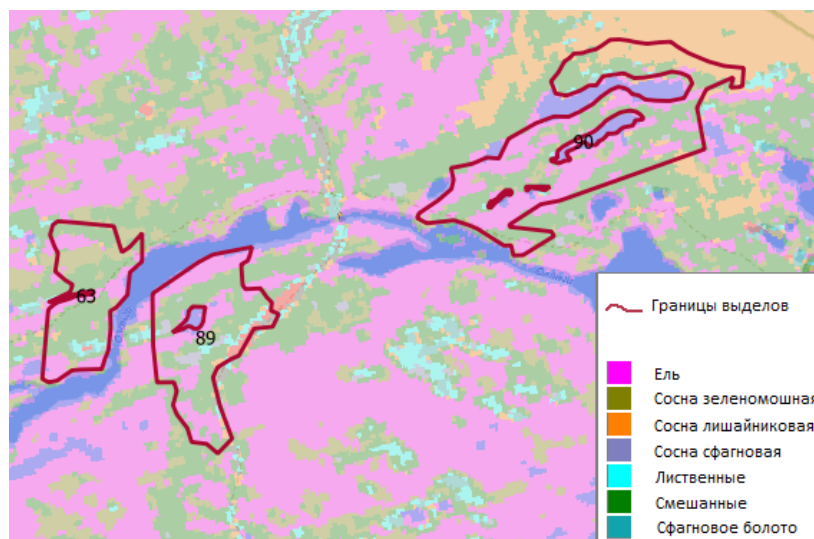


Рис. 4.7. Расположение репрезентативных эталонных участков на фоне слоя растительного покрова.

Эталонные участки нанесены на картографический материал – слой «Растительный покров и антропогенное воздействие». Анализ географической ситуации репрезентативных участков на территории НП «Паанаярви» говорит о том, что растительный покров в их пределах не монотонен, представлены разные породы деревьев, имеются водные объекты, болота и объекты инфраструктуры. Это ставит алгоритмы классификации «в реальные условия».

Для следующего этапа исследования - экспериментальной реализации алгоритмов МО были подобраны безоблачные космические снимки спутников Sentinel-2 и Landsat-8: сцены с разрешением 10-20 м и 30 м, соответственно, за период, предшествующий датам проведения ЛПО [140, 141]. Для Sentinel-2 скачивались снимки уровня обработки 2A, каждый продукт которого состоит из фрагментов площадью 100 км² в картографической проекции UTM/ WGS84 [SENTINEL-2 User Handbook]. Продукт Level-2A предоставляет изображения, в пикселях которых содержатся значения отражательной способности нижней части атмосферы (BOA), полученные из соответствующих продуктов Level-1C, они не требуют атмосферной коррекции.

Загрузку космических снимков в проект можно выполнить двумя путями: импортировать уже заранее скачанные архивом снимки или воспользоваться плагином SCP (см. Гл. 1) и его встроенным инструментом загрузки данных непосредственно через интерфейс плагина. Второй способ имеет некоторые нарекания, поэтому на этапе подготовки обучающей выборки, где важна не автоматизация и скорость, а качественный и репрезентативный материал, использовался импорт извне. Преимуществом загрузки через плагин SCP является предварительный выбор нужных каналов, в отличие от загрузки архивом с сайтов [140, 141].

Для удобства работы целесообразно создать единое изображение, в котором отдельные каналы будут оставаться каждый в своем слое. Далее из такого растрового изображения можно создавать различные композиты для визуального дешифрирования объектов того или иного типа. Для целей исследования нам не нужен весь охват снимка, поэтому мы применяли маску для обрезки (рис. 4.8).

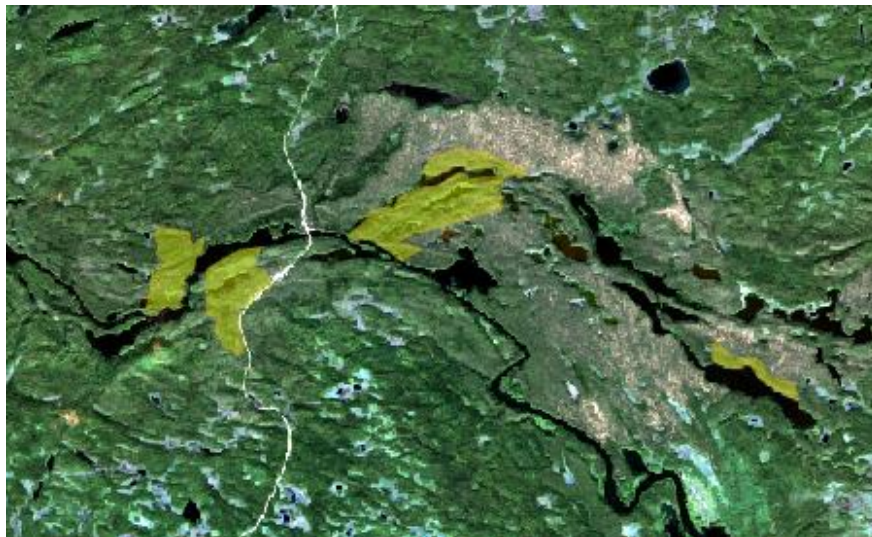


Рис.4.8. Фрагмент снимка Sentinel-2 с нанесёнными контурами эталонных участков. Композит из каналов (B2, B3, B4, B8, B11) с разрешением 10 м. Исходные слои в проекции EPSG:32635 - WGS 84 / UTM zone 35N.

Таким образом в проект загружены все необходимые наборы данных, и завершён базовый этап подготовки обучающих выборок для проведения классификации методами машинного обучения.

4.4.1. Создание обучающей выборки для SAM и классификация

Классификация с обучением может быть выполнена в QGIS с помощью одного из модулей - **SCP** (Semi-Automated Classification Plug-In – плагин для Полуавтоматической Классификации) на основе алгоритма **SAM** (Spectral Angle Mapper - Метод спектрального угла). Тестирование должно быть проведено применительно к участкам, изображения которых соответствующим образом подготовлено (см. 4.4).

Для осуществления полуавтоматической классификации необходимо набрать графические пробы (ROI) отдельных типов покрытий или классов. Акты ЛПО не сообщают конкретного положения поврежденных деревьев в границах выдела, поэтому прежде чем указать класс «повреждение» нам необходимо установить несколько фрагментов поврежденного древостоя самостоятельно. Полигональный слой выделов является ориентиром в данном случае. Чтобы убедиться в том, что древостой на обозреваемой территории действительно не здоров, предварительно осуществляется

сверка с вегетационными индексами NDVI и SWVI. Они рассчитаны при помощи калькулятора растров по формулам (1) и (2). Значения индекса NDVI ниже 0,6 говорят о средней степени развития биомассы (это значение актуально для летнего периода).

$$NDVI = \frac{NIR-RED}{NIR+RED} \quad (1)$$

$$SWVI = \frac{NIR-SWIR}{NIR+SWIR} \quad (2)$$

где $RED \in [0,6 - 0,7]$ мкм, $NIR \in [0,76 - 0,9]$ мкм, $SWIR \in [1,55 - 1,75]$ мкм

Более показательным, особенно в случае ветровалов (для основного исследуемого участка НП «Паанаярви» именно ветра прошлых лет явились причиной образования сухостоя), является индекс SWVI (Shortwave Vegetation Index – коротковолновый вегетационный индекс). Он отражает водный стресс растительности. Фильтрация слоя со значениями индекса $\Delta SWVI$ по пороговому критерию $n \cdot \sigma$ (количество n стандартных отклонений в отдельных исследованиях выбирают исходя из различных соображений), позволяет выявить наиболее существенные изменения, произошедшие в лесном покрове за рассматриваемый временной период [142]. Для выявления ветровалов Шихов А.Н. [143] использовал пороговый критерий разности индекса SWVI ($\Delta SWVI$), равный 1,5 стандартных отклонений, а С.В. Князева - 1 [144].

Для создания поверхности $\Delta SWVI$, где четко выделялись бы последствия ветровалов, нужны снимки до и после стихии. Нам неизвестно, когда именно по территории прошли ветра, повлекшие за собой свал и усыхание древостоя. На данную территорию доступны безоблачные снимки Sentinel-2 за 25 сентября 2018 г. и Landsat-8 за 18 августа 2017 г. По диаграмме распределения спектральной яркости по каналам, приводимой Крыловым А.М. в [99] видно, что наибольшая разница наблюдается на длинах волн 0,85–0,90 мкм, т.е. в 8-м канале снимков Sentinel-2. При использовании синтезов каналов 11-8-4 ветровалы отображаются на снимке в виде светлых полос на фоне салатно-зеленого неповрежденного леса. По синтезам каналов (11-8-4 для Sentinel-2 и 6-5-4 для Landsat-8) мы обнаружили выделяющиеся коричневые фрагменты на фоне зеленой здоровой растительности (рис. 4.9). Мы пришли к выводу, что эти фрагменты и есть локальные ветровальные участки. Создание поверхности индекса $\Delta SWVI$ не дало бы интерпретируемых результатов, поскольку не удалось обнаружить снимков «до».

Выделенные сухостойные участки на снимке относятся к сосновым лесам, что подтверждено сверкой по слою «Растительный покров».

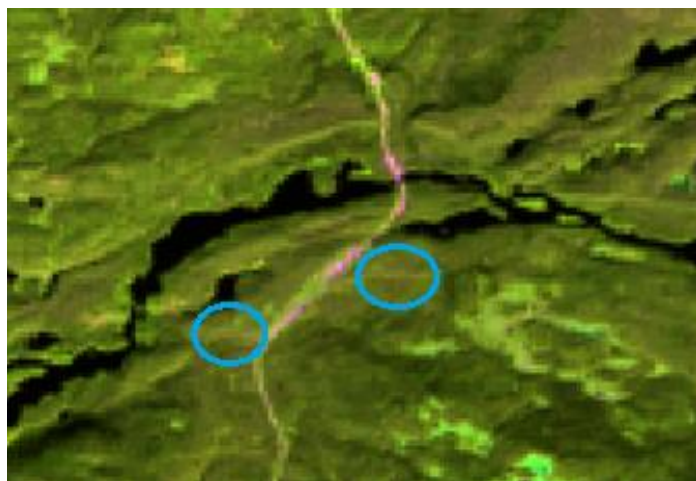


Рис. 4.9. Снимок Landsat-8 за август 2017 г. Комбинация каналов 6-5-4. Небольшие коричневые полосы – ветровалы (обведены синим).

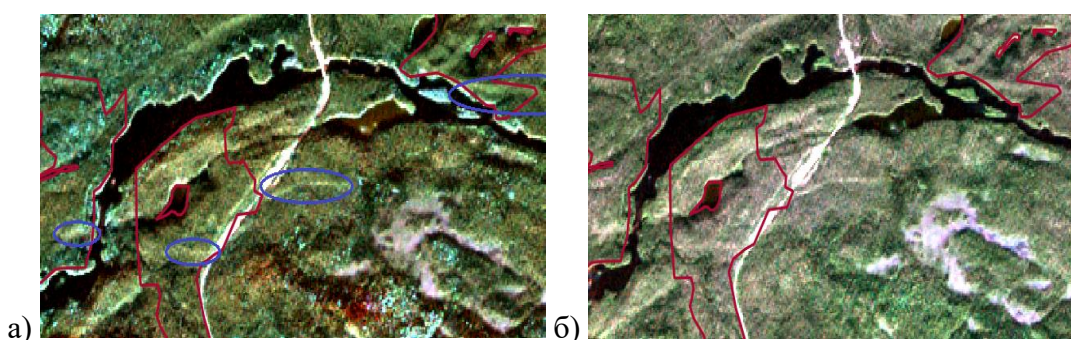


Рис. 4.10. а) Вытянутые объекты на синтезированном изображении (каналы 2-3-4, снимок Sentinel-2 за 25 сентября 2018 г.); б) На синтезированном изображении (каналы 2-3-4, снимок Sentinel-2 за 2 сентября 2019 г.) вытянутые объекты менее различимы. Границы эталонных выделов обозначены красным.

Еще одним этапом сверки было обращение к программе Google Earth Pro и ортофотоплану, созданному на основе снимков сверхвысокого разрешения (рис. 4.11). На данных более высокого разрешения (дата снимка 16 июня 2019 г.) можно отделить безлесные пространства от ветровалов, схожих по цвету и тону.

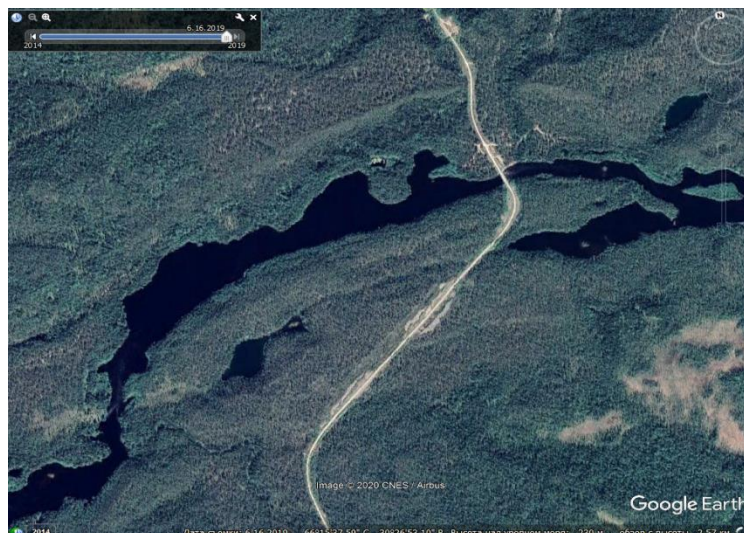


Рис. 4.11. Фрагмент снимка из программы Google Earth Pro.

Идентифицировав участки целевого класса – поврежденных лесов, мы создали обучающую выборку (набрали графические пробы инструментами плагина SCP).

Классификация проводилась по 7 классам (рис. 4.12). В результате классификации с помощью алгоритма было получено растровое одноканальное изображение, где каждый пиксель содержит значение от 0 до 7.

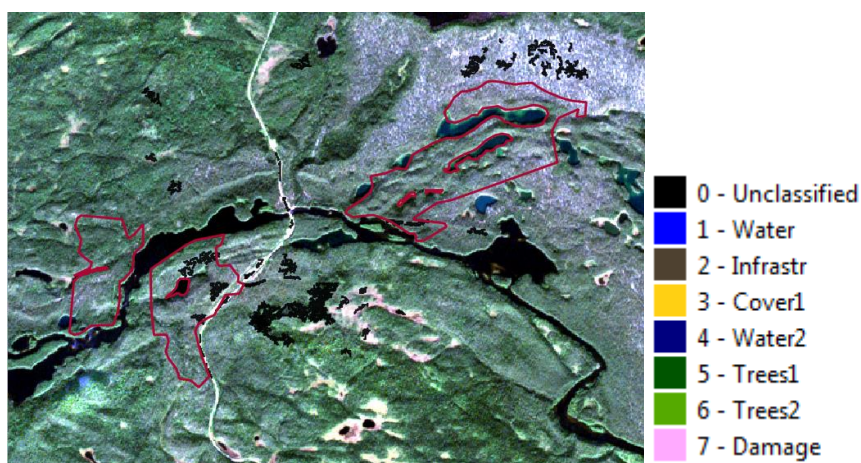


Рис. 4.12. Распространение графических проб по территории исследования (черные «пятна») и выделяемые классы.

Визуальная оценка схожести полуавтоматической классификации алгоритмом SAM с покрытием «Растительный покров и антропогенные ландшафты РК» (данные на 2011 г. можно считать приемлемыми в плане актуальности если не для инфраструктуры и нарушений, то для типа растительности за счет высокой долговременности смены сосново-еловых пород) показывает хороший результат.

Инструмент «шторка» MapSwipe Tool (рис. 4.13) применен для повышения наглядности классификации, так как другие способы - умножение, наложение (рис. 4.14), задание полупрозрачности – более трудоемкие. Однако, эффективность и этого

инструмента не абсолютна. Так, например, классификатор определил водоемы в тех местах, где визуально мы опознаем опушку; на слое «Растительный покров и антропогенные ландшафты» эти водоемы присутствуют, то есть алгоритм прав. Классификатор определяет воду с высокой точностью, потому что, вручную создавая обучающую выборку, мы постарались учесть разные типы водных поверхностей (синие, мутно-синие, коричневого оттенка). Иная ситуация складывается с определением участков преобладающих пород - ельника, сосняка, смешанного леса - из-за низкой репрезентативности этих классов в обучающей выборке. Значения целевого класса – повреждения – классификатор присвоил массивам лишайниковой сосны, расположенным на северо-востоке от выделов, на информацию с которых мы опирались при создании обучающей выборки. Подтвердить или опровергнуть результаты этого определения мы не можем, поскольку, несмотря на проведение визуального дешифрирования несколькими методами, мы не знаем точного расположения погибших деревьев в границах эталонного выдела.

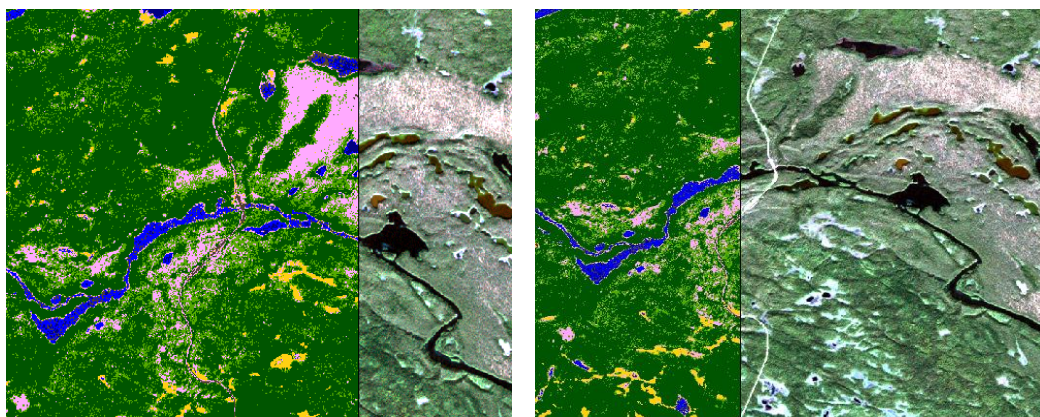


Рис. 4.13. Сопоставление классифицированного изображения с космическим снимком

Высокая способность классификатора идентифицировать пиксели из других классов (вода, инфраструктура) по их спектральной схожести даёт основания полагать, что для создания более достоверной обучающей выборки и улучшения работы классификатора нам необходимо однозначно знать и указать пиксели с сухостоем, что маловероятно.

Чтобы создать читаемое картографическое изображение, необходимо произвести исключение (отсеивание) мелких растровых деталей. Точность, с которой можно проводить исключение мы оценивали, опираясь на «Порядок проведения ЛПМ...» (см. Гл. 3). При разрешении исходных аэрокосмических данных 10 м, надо фиксировать 1 га. Отсюда, квадрат размером 10×10 пикселей является площадным порогом для исключения. Минусом такого подхода является то, что объекты инфраструктуры (например, по

территории НП «Паанаярви» в области исследования проходит дорога, линейный объект, ширина которого в разных местах 1-3 пикселя – ширина дорожного полотна не более 10 м, вместе с обочиной - не более 20 м) частично замещаются более крупными соседними группами. Но в рамках задачи лесопатологического мониторинга этим фактом можно пренебречь. Для выявления же более мелких контуров повреждений может возникнуть необходимость приобретения снимков сверхвысокого разрешения.

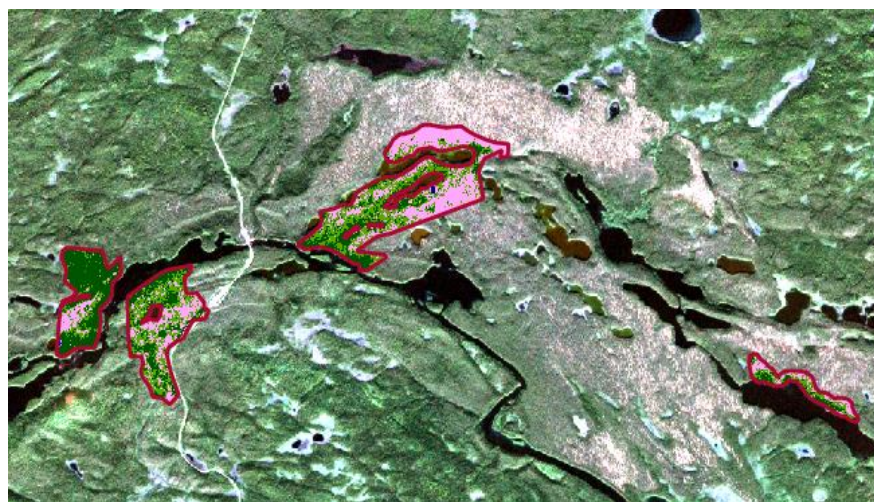


Рис. 4.14. Наложение только классифицированных эталонных выделов на космический снимок.

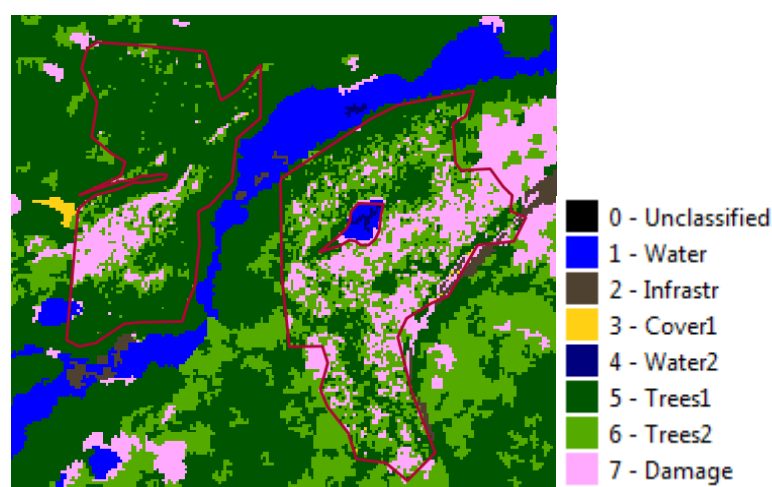


Рис. 4.15. Классифицированное изображение, преобразованное сглаживанием и классифицированные изображения выделов (внутри красных контуров) без фильтрации.

Мы произвели сглаживание сигнала в ячейках растра и «отсеяли» шум. Итоговое растровое изображение пригодно для векторизации (рис. 4.15). Таким образом была выполнена классификация при помощи алгоритма SAM.

4.4.2. Создание обучающей выборки для Random Forest

В ходе обзора литературы было отмечено, что ансамблевый алгоритм RF зарекомендовал себя для работы с ДДЗЗ (см. Гл.2., п.2.4). Нами был выбран данный алгоритм для классификации покрова бореальных лесов с нарушениями в них. Код модели был нами специально написан на высокоуровневом языке программирования Python в среде Google Colaboratory.

Фрагменты сухостоя по эталонным участкам были определены в ходе подготовки данных для классификации алгоритмом SAM. Для классификации Random Forest было решено использовать каналы 2, 3, 4, 8 и 11. В процессе эмпирического поиска было замечено, что комбинация 8-11-4 позволяет отличать темнохвойные леса от светлохвойных весьма отчетливо. При этом лишайниковая сосна также отличима.

На рис. 4.16. приведена схема подготовки данных для создания набора данных в среде Python: из композита вырезан прямоугольный фрагмент, содержащий один эталонный выдел (1); на основе этого фрагмента создано 5 одноканальных изображений, соответствующих 5 классам (2); средствами библиотек osgeo, pandas, numpy каналы мультиспектрального снимка и изображения по классам загружены в таблицу данных (3-4). Полный код модели – в приложении настоящей ВКР.

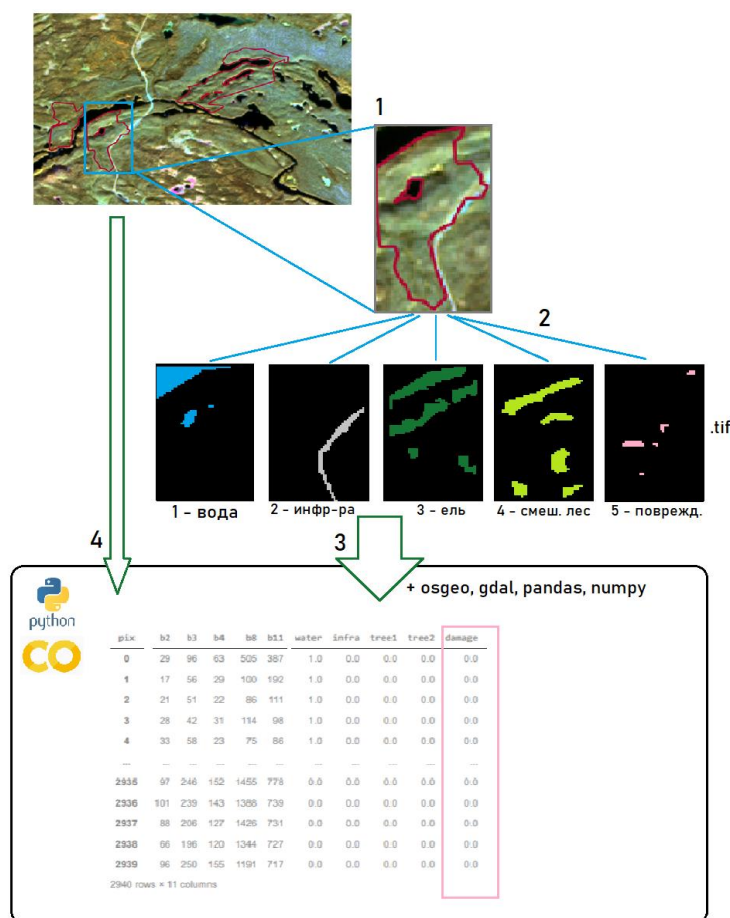


Рис. 4.16. Схема подготовки обучающей выборки для модели RF в Python.

Здесь надо отметить, что при многоклассовой классификации важно соблюсти равенство экземпляров (в нашем случае пикселей) каждого класса, чтобы при обучении не произошел «скос» в сторону количественно доминирующего класса. В создании выборки мы опирались на процент поврежденного древостоя, как на общий знаменатель и для остальных классов. Из обучающей выборки были затем удалены все экземпляры, имеющие значение «0» в признаке класса («class»).

Для конструирования признаков были проведены нормализация и анализ числовых данных для выявления тенденций. Поскольку данные каналов видимого диапазона (B2, B3, B4) коррелируют между собой, был проведен анализ главных компонент (PCA), и количество признаков сокращено до 3-х при общем объеме выборки 1295 экземпляров.

4.4.3. Создание модели Random Forest и классификация

Исходный набор данных был разбит на тренировочный и тестовый поднаборы в пропорции 3:2 (777:518). Созданная модель была обучена и протестирована. Точность (accuracy) составила около 81 % (рис. 4.17).

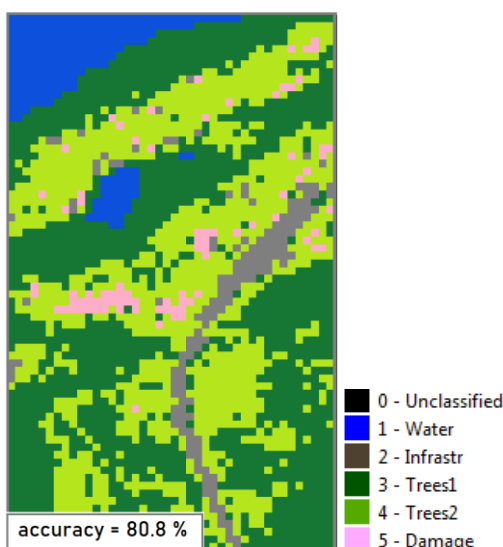


Рис. 4.17. Результат классификации RF исходного набора данных.

В ходе построения модели Random Forest нам пришлось сократить объем обучающей выборки, удалив экземпляры, для которых известны значения каналов космического снимка, но которые не попали в подготовленные нам фрагменты классов. Для тренировочного набора данных было взято 60 % обучающей выборки, использовано 100 Деревьев решений.

4.5. Оценка работоспособности модели МО

Поскольку целью исследования является именно детектирование верно-определенных участков с поврежденным древостоем на эталонных выделах, то не имеет смысл рассчитывать точность классификации для большей территории. В итоге предложена следующая последовательность оценки классификации методами МО:

1. «обрезать» классифицированное изображение по векторному слою эталонных выделов (для них известен процент поврежденного древостоя) и сохранить в отдельные файлы изображений;
2. рассчитать долю (и процент) пикселей со значением искомого класса (7 и 5 для SAM и RF, соответственно).


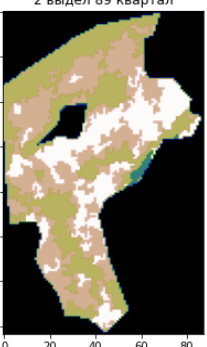
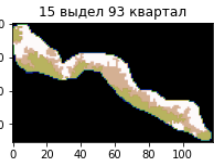
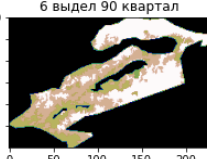
$$\delta = \frac{M_8}{N} * 100 \quad (3),$$

где M_8 — количество пикселей со значением 8, N - количество всех пикселей.

Изображение каждого выдела сохраняется в виде прямоугольной матрицы, при этом пиксели, находящиеся за границей выдела, но в пределах прямоугольника изображения хранят числовые значения, не имеющие отношения к классам. Чтобы в знаменателе формулы (3) учитывались только значимые пиксели, на Python нами был написан скрипт. Он подсчитывает процент пикселей с повреждённым лесом (по мнению модели), что затем можно сравнить с данными из таблицы 4.1. «Характеристики эталонных территорий с зафиксированными повреждениями», составленной по актам ЛПО. Это аналогия метрик классификации.

Результаты применения алгоритма SAM приведены в табл. 4.2.

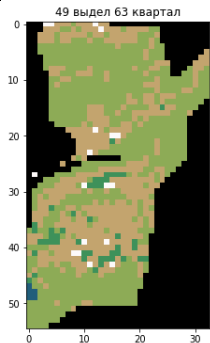
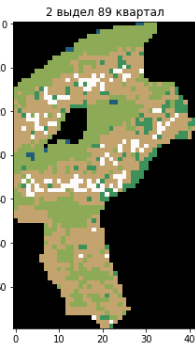

Табл. 4.2. Количественные показатели работы алгоритма SAM.

	49 выд. 63 кв.	2 выд. 89 кв.	15 выд. 93 кв.	6 выд. 90 кв.
0 - Unclassified 1 - Water 2 - Infrastr 3 - Cover1 4 - Water2 5 - Trees1 6 - Trees2 7 - Damage				
Доля каждого класса внутри выдела, %	7: 11.56 5: 77.77 6: 10.19	7: 25.12 5: 35.36 6: 38.50	7: 41.27 5: 29.31 6: 28.99	7: 45.27 5: 20.01 6: 33.97

	1: 0.49	1: 0.13 2: 0.89	1: 0.42	4: 0.68 1: 0.07
Поврежденный древостой по ЛПО, %	23	13	10.5	10.5
Относительная погрешность модели	0.496	0.931	2.933	3.314

Результаты применения алгоритма Random Forest приведены в табл. 4.3.

Табл. 4.3. Количественные показатели работы алгоритма RF.

	49 выд. 63 кв.	2 выд. 89 кв.	6 выд. 90 кв.
0 - Unclassified 1 - Water 2 - Infrastr 3 - Trees1 4 - Trees2 5 - Damage			
Доля каждого класса внутри выдела, %	5: 1.53 4: 36.65 3: 57.48 2: 3.91 1: 0.43	5: 7.85 1: 0.76 3: 35.91 4: 47.50 2: 7.98	5: 7.31 3: 45.26 4: 29.09 2: 17.77 1: 0.57
Поврежденный древостой по ЛПО, %	23	13	10,5
Относительная погрешность модели	0.935	0.392	0.305

Из приведенных выше количественных показателей работы классификаторов видно, что модель на основе ансамблевого алгоритма RF в целом дала более близкие к контрольным данным результаты. Очень по разному классификаторы повели себя в отношении выдела 6 в квартале 90 – том, где присутствуют массивы лишайниковой сосны.

Визуализация выявления повреждений древостоя методами МО.

Открытая ГИС QGIS позволяет полученную производную информацию визуализировать – представить в виде полноценных картографических изображений. Классифицированные изображения были преобразованы в карты с легендами (рис. 4.18, 4.19). Изображение, сгенерированное непосредственно в среде разработки, было затем дополнено векторным слоем с границами выделов из проекта QGIS (рис. 4.20).

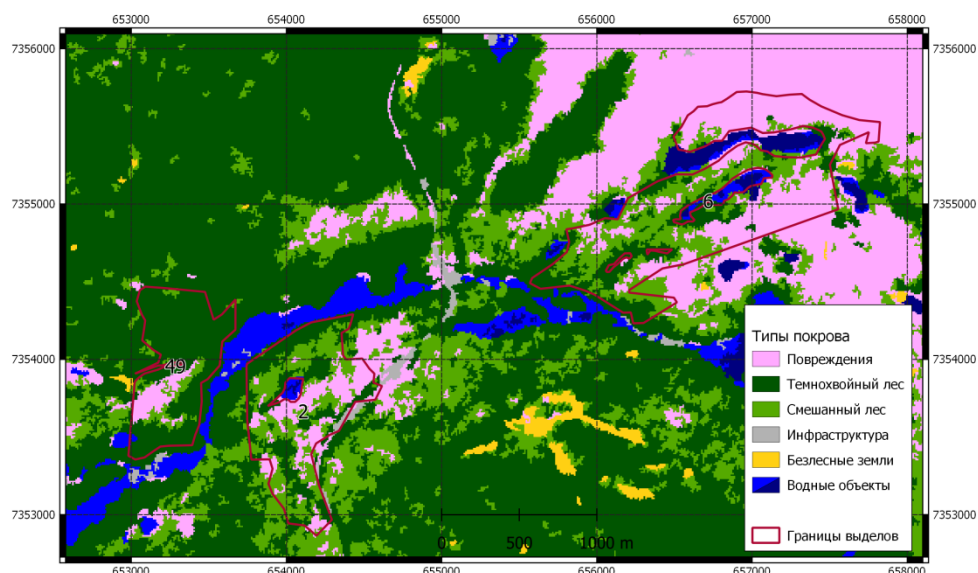


Рис. 4.18. Картографическое изображение «Типы растительного покрова с деградациями древостоя на основе классификации с помощью алгоритма SAM».

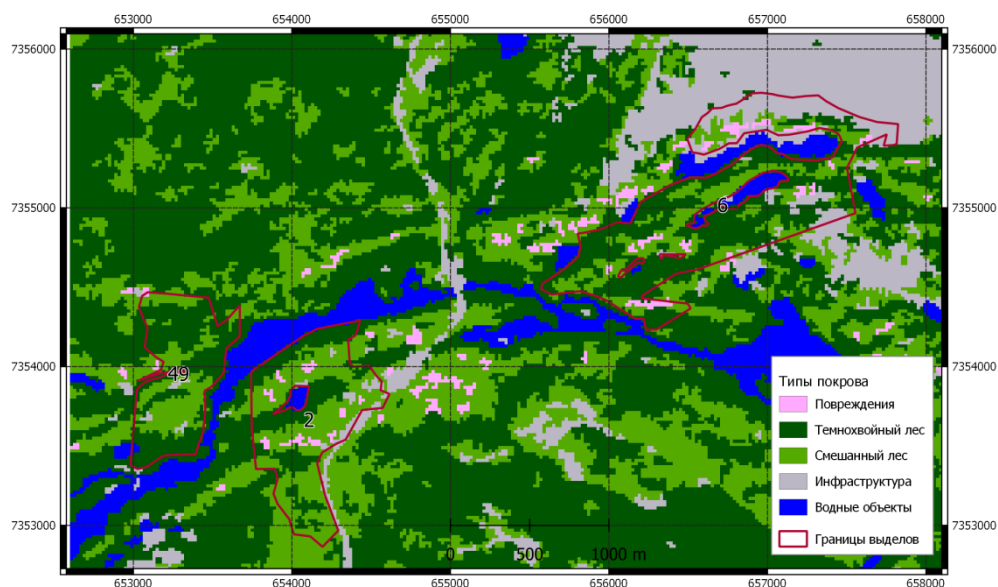


Рис. 4.19. Растровая карта «Типы растительного покрова с деградациями древостоя на основе классификации с помощью алгоритма RF».

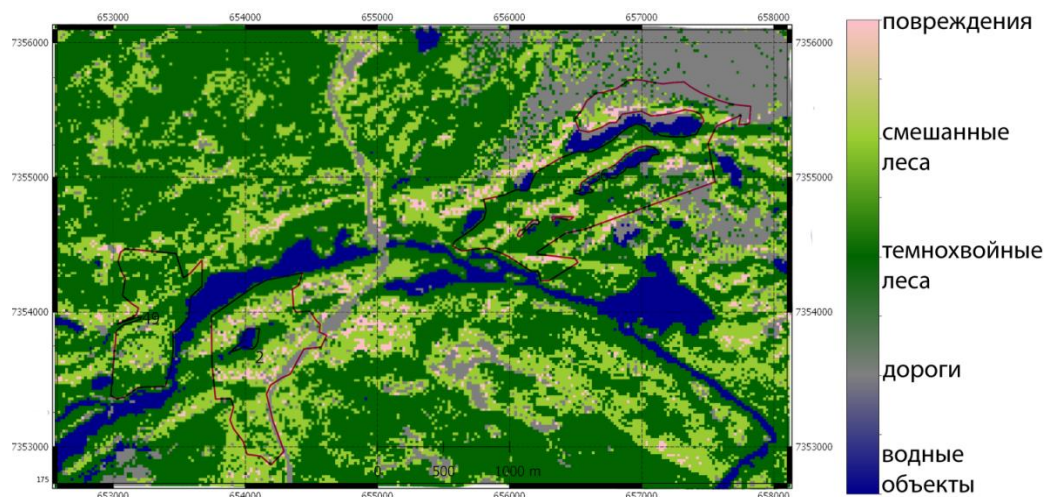


Рис. 4.20. Растровая карта «Типы растительного покрова с деградациями древостоя на основе классификации с помощью алгоритма RF», сгенерированная непосредственно в среде Python.

По результатам проведения автоматической классификации космических снимков при помощи алгоритмов машинного обучения составлены карты типов растительного покрова с деградациями древостоя в масштабе 1:25000.

4.6. Обсуждение выявления повреждений древостоя методами МО

Проведенное экспериментальное распознавание повреждений лесов по снимкам Sentinel-2 2A с применением двух алгоритмов МО позволяет сделать некоторые выводы:

- материалы актов ЛПО не являются оптимальными исходными данными для формирования эталонных экземпляров. Лесоустроительные выделы обладают высокой пространственной неоднородностью, что делает визуальное выявление повреждений без привлечения экспертного мнения очень ненадежным. Некорректный выбор экземпляров целевого класса может приводить к ошибкам;

- алгоритм **SAM** с высокой долей точности определяет крупные формы ландшафта и большие однородные объекты, при наличии репрезентативных эталонных данных. На исследуемой территории были идентифицированы объекты инфраструктуры и водоемы. Для визуального отбора эталонов были использованы синтезированные изображения разных каналов, в том числе разновременные композиты SWVI, и были выбраны соответствующие пиксели. Тем не менее, данный алгоритм не смог разделить спектральные значения отобранных каналов должным образом. На примере классификации выделов 15/93 и 6/90 видно, что класс 7 (повреждение) был приписан участкам произрастания лишайниковой сосны, где сомкнутость полога очень низкая.

Ошибочное отнесение к ветровальным сухостоям таких нелесных участков, как прогалины и опушки можно избежать наложением масок, созданным по сценам, снятым до ветровала. Для этого необходимо обладать метеорологическими сводками на территорию за достаточно длительный период. Главное отличие сухостоя, образованного на ветровале от таких участков – это их вытянутая форма. Алгоритм SAM осуществляет классификацию исключительно по значениям в пикселях, и в целом для определения фрагментарных повреждений древесных насаждений на повыведельном уровне оказался недееспособным;

- ансамбль из Деревьев решений **RF** показал противоположный результат, занижив показатели поврежденного леса на тестовых участках. Одним из объяснений этому мы видим недостаточную сбалансированность классов в обучающей выборке. Пространственная ориентация участков, отнесенных моделью в категорию поврежденных, заставляет положительно отзываться о качестве выявления местоположений ветровалов. Большинство детектированных моделью участков представляют собой полосы, вытянутые в северо-восточном направлении, что соответствует типичному направлению ветров, дующих в этом регионе в летний период. То есть, в отличие от алгоритма SAM, модель на основе RF оказалась способна фиксировать пространственную картину процесса повреждений леса в той или иной мере.

Пути развития идеи применения МО при лесопатологическом мониторинге.

В Главе 2 говорится, что и пожары, и ветровалы, и рубки имеют разработанные методы детектирования, но мы не встретили работы, инкорпорирующей распознавание перечисленных типов нарушений (гарь, ветровал, болезнь, рубка), как и работы, использующие Python библиотеку eo-learn для этих целей. Нами была оценена применимость алгоритмов МО в задачах выявления нарушенного древостоя безотносительно его происхождения. Несмотря на то, что на данный момент модели продемонстрировали качество классификации недостаточное для их ввода в действие, очевидно, есть пути для улучшения их работы. В первую очередь, необходимо привлечение качественных и современных данных полевых исследований для повышения достоверности эталонных данных, также нужно продолжить поиск оптимальных гиперпараметров модели в случае с RF (количество деревьев решений, глубина и др.). Поскольку ансамбль Деревьев решений продемонстрировал более приближенные к контрольным данным результаты, то нам видится перспективным апробация других ансамблевых алгоритмов. В частности, алгоритма LightGBM реализованного в специализированной Python-библиотеке eo-learn.

Развитие данной темы мы видим также в решении задачи по выявлению по снимкам причин отсутствия леса: если нарушения лесного покрова выявляются моделью с высокой степенью достоверности, то можно предпринять попытку классификации именно типов повреждения леса. Для этого необходимо понимание механизмов образования каждого конкретного типа повреждений или нарушений лесов. Решение данной задачи может способствовать оценке количества ресурсов лесной экосистемы. Динамика изменений биомассы и продуктивности, происходящих в ходе сукцессий зависит от вида нарушений и типа системы в первую очередь, поэтому для оценки ресурсов лесной экосистемы важно понимать, чем был вызван урон лесов на конкретной территории. Из этого следует, что знание типов нарушения лесов может повысить качество прогнозирования лесной динамики.

Заключение

Обстоятельное знакомство с методами исследования, оценки и картографирования природных ресурсов позволило выделить актуальное направление – применение данных дистанционного зондирования совместно с машинным обучением в их обработке. Нам представилось возможным рассмотреть сущность и возможности МО, как реализации искусственного интеллекта, его разных алгоритмов в контексте использования пространственных данных средствами ГИС (Глава 1) в выявлении нарушений леса. Для этого рассмотрены многочисленные аспекты лесной отрасли – экологическая и экономическая значимость леса, его функции, управление, инвентаризация, мониторинг, основные причины нарушений или отсутствия древостоя на лесных территориях, в том числе в зоне бореальных лесов северо-западной части России, имеющей географические сходства с северо-европейскими зарубежными лесами, где ведение лесного хозяйства считается успешным (Глава 2).

Анализ опыта применения МО в лесном секторе показал, что чаще всего выбор исследователей падает на алгоритмы k-NN, SVM, Деревья решений и ансамбль Деревьев решений – Random Forest, а также глубокое обучение. В ходе фокусирования на состоянии лесной промышленности и системах учета лесов Карелии (в пределах особо охраняемых территорий) обнаружилось, что необходимость получения оперативной информации о состоянии лесов в регионе трактуется не только общими разумными соображениями, но и официальным переходом на другую модель ведения лесного хозяйства (Глава 3). Показано, какие методы обработки открытых спутниковых изображений позволяют идентифицировать сухостой, образовавшиеся на ветровалах, и какие части спектра являются наиболее репрезентативными.

В основу набора данных о повреждениях древостоя легли материалы спутниковой съемки Sentinel-2 и Landsat-8, документальные акты лесопатологических обследований и тематические карты растительного покрова. В исследовании тестировали два принципиально разных алгоритма – Spectral Angle Mapper (SAM, реализованный в плагине SCP под QGIS) и Random Forest (RF), модель на основе которого была написана нами на языке Python. Для сравнения качества классификации лесного покрова двумя моделями был разработан и написан также на Python способ оценки полученных результатов. Рассчитанная относительная погрешность детектирования «сухостойных» пикселей оказалась очень большой для алгоритма SAM. Модель RF, показавшая точность 81 %, сработала лучше в отношении выбранной территории. Для обоих алгоритмов «камнем преткновения» стали массивы лишайниковой сосны, под которой слишком хорошо видна поверхность. Из этого следует, что в данном случае необходимо либо

объединить похожие категории растительного покрова в один класс, либо активно работать над разделением чрезвычайно похожих классов путем увеличения размера их обучающей выборки.

Исходя из масштабного ряда и детальности растровых карт, продиктованными нормативно-правовой базой, было оценено, насколько применение космических снимков Sentinel-2 вкупе с точностью работы самой модели МО, применимо для выявления нарушенного древостоя. Результаты классификации по алгоритму SAM можно считать удовлетворительными с точки зрения общего распознавания типов покрытий для ускорения картографирования территории, но на уровне лесных выделов он видится малопригодным. По результатам работы модели RF удалось составить карту в масштабе 1:25 000, которая может служить основанием для инициирования полевых инструментальных обследований проблемных участков. Также, полученное изображение при совместном анализе с другими информационными источниками (карт породного состава леса и запасов) может быть полезно для ориентировочной оценки потерь древесины или ущерба углеродному балансу.

Оценивая применимость алгоритмов машинного обучения для картографирования лесов надо сказать, что точность создаваемых карт зависит от качества и объема обучающей выборки. На снимках с разрешением 10 м спектральная информация одного пикселя это сложная комбинация излучения, отраженного от нескольких деревьев, возможно разных пород, и от подстилающей поверхности. Это, и тот факт, что для подготовки контрольных данных были использованы акты ЛПО и лесоустроительные материалы, где показатели приводятся на неоднородные по своему составу выделы, говорят о том, что обучающие данные сложно назвать оптимальными. Несмотря на это, модель RF позволяет создавать интерпретируемое картографическое изображение.

Таким образом в исследовании все поставленные задачи выполнены. Полученные результаты дают основания полагать, что, обучая ансамблевые алгоритмы на данных более высокого пространственного разрешения и выделяя экземпляры целевого класса с большей достоверностью можно добиться более точной классификации. Одним из идейных продолжений работы видится классификация лесных массивов по отдельным типам повреждений с применением ансамблевого алгоритма LightGBM.

Список использованных источников

1. Потапов, А. С. Искусственный интеллект и универсальное мышление. – СПб. : Политехника, 2012. — 711 с.
2. MOOC Artificial Intelligence (Accenture) // URL: <https://www.futurelearn.com/courses/artificial-intelligence>. Дата обращения: 21.01.2020
3. Austin P. C. et al. Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes // Journal of clinical epidemiology. – 2013. – Т. 66. – № 4. – С. 398-407.
4. Bassier M., Van Genechten B., Vergauwen M. Classification of sensor independent point cloud data of building objects using random forests // Journal of Building Engineering. – 2019. – Т. 21. – С. 468-477.
5. Thessen A. (2016) Adoption of Machine Learning Techniques in Ecology and Earth Science.
6. Смирнова Ю. Маленький агент глобального потепления // Журнал СПбГУ № 2 (3910), март 15, 2018.
7. AWS powers F1 Insights // URL: <https://aws.amazon.com/f1/>. Дата обращения: 22.04.2020
8. URL: https://vas3k.ru/blog/machine_learning/. Дата обращения: 01.11.2019
9. Коэльо, Л.П., Ричарт, В. Построение систем машинного обучения на языке Python. 2-е издание / пер. с англ. Слинкин А. А. – М. : ДМК Пресс, 2016. – 302 с.
10. Andreo, V., Belgiu, M. Rodents and satellites. (2019). Predicting mice abundance and distribution with Sentinel-2 data. // Ecol. Informatics, 51, 157-167.
11. Csillik, O., Belgiu, M. (2019). Object-Based Time-Constrained Dynamic Time Warping Classification of Crops Using Sentinel-2. // Remote Sens. 11, 1257.
12. Модель зависимости // URL: http://www.machinelearning.ru/wiki/index.php?title=Модель_алгоритмов. Дата обращения: 22.02.2020
13. Классический набор данных недвижимости в районах Бостона Boston Housing dataset // URL: <https://www.datasciencecentral.com/profiles/blogs/boston-housing-dataset-without-the-racial-profiling-field> Дата обращения: 22.02.2020
14. Kobler A. & Adamič M. Identifying brown bear habitat by a combined GIS and machine learning method // Ecological Modelling. 135(2):291-300, December 2000.
15. URL: https://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задачах_классификации_и_регрессии Дата обращения: 19.01.2020
16. Дьяконов, А. Функционалы качества бинарной классификации // URL:<https://dyakonov.org/> Дата обращения: 22.02.2020

17. URL: <http://datareview.info/article/luchshaya-metrika-dlya-ocenki-tochnosti-klassifikacionnykh-modelej/> Дата обращения: 23.02.2020
18. Метрики в задачах машинного обучения
URL: <https://habr.com/ru/company/ods/blog/328372/> Дата обращения: 23.02.2020
19. Naghibi, S. & Pourghasemi, H. R. & Dixon, B. (2016). GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. // *Environmental Monitoring and Assessment*. 188.1 (2016): 44.
20. Mollalo, A., Sadeghian, A., Israel, G.D., Rashidi, P., Sofizadeh, Glass, G. E. (2018). Machine learning approaches in GIS-based ecological modeling of the sand fly *Phlebotomus papatasi*, a vector of zoonotic cutaneous leishmaniasis in Golestan province, Iran. // *Acta tropica*. – 2018. – Т. 188. – С. 187-194.
21. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D. (2018). Machine Learning in Agriculture: // *Sensors*. – 2018. – Т. 18. – №. 8. – С. 2674
22. Bagheri, M. & Bazvand, A. & Ehteshami, M. (2017). Application of artificial intelligence for the management of landfill leachate penetration into groundwater, and assessment of its environmental impacts // *Journal of Cleaner Production*. 149.
23. Черниковский, Д. М., Алексеев, А. С. Метод определения характеристик лесов на основе материалов дистанционного зондирования Земли, данных лесоустройства и алгоритма k-NN (на примере Лодейнопольского лесничества Ленинградской области) // *Известия высших учебных заведений. Лесной журнал*. – 2019. – №. 4 (370).
24. Zhou, B. et al. (2004). Automated soil resources mapping based on decision tree and Bayesian predictive modeling. // *J. Zhejiang Univ. Sci. A* 5, P. 782–795.
25. Mahoney, C. et al. Forest Attribute Mapping Framework: A Pilot Study in a Northern Boreal Forest, Northwest Territories, Canada // *Remote sensing*–2018–Т.10–№.9–С. 1338.
26. Введение в машинное обучение // URL: <https://habr.com/ru/post/448892/> Дата обращения: 3.09.2019
27. Классификатор kNN // URL: <https://habr.com/ru/post/149693/> Дата обращения: 2.09.2019
28. Замай С. С., Охонин В. А., Якубайлик О. Э. Нейронные сети и ГИС // *Основы геоинформатики*. – 2004. – №. 2. – С. 255-266.
29. URL: <https://www.quora.com/What-are-applications-of-machine-learning-in-Geographic-Information-Systems-GIS> Дата обращения: 3.09.2019
30. URL: <https://medium.com/sentinel-hub/introducing-eo-learn-ab37f2869f5c> Дата обращения: 12.03.2020

31. Use cases: Oil Inventories // URL: <https://orbitalinsight.com/use-cases/oil-inventories> (20.04.2019)
32. Accenture и Intel спасают коралловые рифы с помощью искусственного интеллекта / URL: <https://www.accenture.com/ru-ru/about/company/company-news-release-accenture-intel-save-coral-reefs> Дата обращения: 07.04.2020
33. Artificial Intelligence Helps Distinguish the Forest From the Trees // URL: <https://blog.globalforestwatch.org/data-and-research/artificial-intelligence-helps-distinguish-the-forest-from-the-trees-part-1> Дата обращения: 07.04.2020
34. URL: <https://medium.com/kaggle-blog/planet-understanding-the-amazon-from-space-1st-place-winners-interview-bf66fb444bc2> Дата обращения: 07.04.2020
35. URL: <https://www.kaggle.com/c/planet-understanding-the-amazon-from-space> Дата обращения: 07.04.2020
36. Aksoy, S. et al. (2009). Land Cover Classification with Multi-Sensor Fusion of Partly Missing Data // Photogrammetric Engineering & Remote Sensing. – 2009. – Т. 75. – №. 5. – С. 577-593.
37. Rodriguez-Galiano V. F., Chica-Rivas M. Evaluation of different machine learning methods for land cover mapping of a Mediterranean area using multi-seasonal Landsat images and Digital Terrain Models // International Journal of Digital Earth. – 2014. – Т. 7. – №. 6. – С. 492-509.
38. Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2017). Convolutional neural networks for large-scale remote-sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 55(2), 645-657.
39. Попов, С. Ю. Геоботаническая карта Пинежского заповедника // Геоботаническое картографирование. – Санкт-Петербург, 2018. – С. 3–18.
40. URL: <https://pro.arcgis.com/ru/pro-app/help/analysis/image-analyst/the-image-classification-wizard.htm> Дата обращения: 17.01.2020
41. Congedo Luca (2016). Semi-Automatic Classification Plugin Documentation. DOI: <http://dx.doi.org/10.13140/RG.2.2.29474.02242/1> Дата обращения: 17.01.2020
42. URL: <https://docs.scipy.org/doc/scipy/reference/> Дата обращения: 23.10.2019
43. Документация по EO-learn // URL: <https://eo-learn.readthedocs.io/en/latest/index.html> Дата обращения: 22.10.2019
44. Классификация покрова земли при помощи eo-learn. Часть 1 // URL: <https://habr.com/ru/post/452284/> Дата обращения: 22.10.2019
45. ФЗ «Об охране окружающей среды» [от 10.01.2002 N 7-ФЗ (ред. от 27.12.2019)] – URL: <http://docs.cntd.ru/document/901808297> Дата обращения: 22.02.2020

46. Основы устойчивого лесопользования: учеб. пособие для вузов / М. Л. Карпачевский; Всемирный фонд дикой природы (WWF). — М., 2009. — 143 с.
47. Антропьянская, Л.Н. Программа «Совместная реализация курса «Основы проектирования» / Томск, 2018. — URL: <https://ppt-online.org/353239>
48. Егоренков, Л.И. Статистика природопользования: Уч. пособие / Егоренков Л.И. — М. : Форум, НИЦ ИНФРА-М, 2015. — 176 с.
49. Сизов, А.П., Антонова, О.М. Анализ расходов федерального бюджета (2016) на деятельность в целях адаптации окружающей среды под нужды человека // Известия вузов. Геодезия и аэрофотосъемка. Сборник статей по итогам научно-технической конф. — 2018. — С. 161-166.
50. Уткин, А.И., Замолодчиков, Д.Г. Углеродные пулы фитомассы, почв и депонирование углерода в еловых лесах России // Хвойные бореальной зоны. — 2004. Вып. 2. — С. 21-30.
51. Уткин, А.И., Замолодчиков, Д.Г., Честных, О.В. Пулы углерода фитомассы и почв сосновых лесов России // Хвойные бореальной зоны. — 2004. Вып. 2. — С. 13-21.
52. Сунгурова, Н.Р., Худяков, В.В., Страздаускас, С.Е. Сравнительная структура углеродного пула в надземной фитомассе культур сосны и ели // Известия высших учебных заведений. Лесной журнал. — 2019. — №. 3 (369).
53. Hu T. et al. Mapping global forest aboveground biomass with spaceborne LiDAR, optical imagery, and forest inventory data // Remote Sensing. — 2016. — Т. 8. — №. 7. — С. 565.
54. Груздев, В. С. Биоиндикация состояния окружающей среды. Монография / В. С. Груздев. — ИНФРА-М, 2018. — 160 с.
55. Шилов, Н., Боровлев, А. Оценка потенциальных рисков для популяций атлантического лосося в случае сокращения площади нерестоохраняемых полос лесов в Северном рыбохозяйственном бассейне // Устойчивое лесопользование. — 2019. — №. 2. — С. 5-13.
56. Angelstam P. et al. Получение новых знаний и обучение в сфере устойчивых лесных ландшафтов: запад и восток Европы как лаборатория // Лесной журнал. — 2019. — №. 1. — С. 9-31-9-31.
57. Global Forest Watch (GFW) URL: <https://www.globalforestwatch.org/> Дата обращения: 8.04.2019
58. Лесной Кодекс Российской Федерации. от 04.12.2006 N 200-ФЗ (ред. от 24.04.2020) — URL: http://www.consultant.ru/document/cons_doc_LAW_64299/ Дата обращения: 22.05.2020

59. Площадь лесных земель // URL: https://www.lesonline.ru/les_squares/ Дата обращения: 15.05.2020
60. Земельный Кодекс Российской Федерации. от 25.10.2001 N 136-ФЗ (ред. от 18.03.2020). Дата обращения: 22.05.2020
61. Страхов, В.В. Глобализация лесного хозяйства. – М. : ВНИИЦлесреурс, 2001. – 400 с.
62. Intergovernmental Panel on Climate Change (IPCC) Formulation of Response Strategy // Geneva, World Meteorological Organization, 1990.
63. Анисимова, П. С. Особенности лесоустройства особо охраняемых природных территорий // Лесной вестник / М.: ГОУ ВПО МГУЛ. – 2009. – № 4 (67). – С. 42–44.
64. Площадь лесных земель // URL: <http://rosleshoz.gov.ru/opendata/7705598840-ForestlandArea> Дата обращения: 17.02.2020
65. Общий запас древесины лесных насаждений на землях лесного фонда и землях иных категорий // URL: <http://rosleshoz.gov.ru/opendata/7705598840-WoodStock> Дата обращения: 17.02.2020
66. Экономическая оценка лесных ресурсов России 16.07.2016 // Лесная промышленность: [сайт]. URL: http://wood-prom.ru/analitika/14650_ekonomicheskaya-otsenka-lesnykh-resursov-rossii Дата обращения: 27.01.2020
67. Писаренко, А.И. Лесное хозяйство России. От использования к управлению. – М. : ИД «Юриспруденция», 2004. – 552с.
68. История русского леса // ЛесПромИнформ. – 2006. – № 6 (37). // URL: <https://lesprominform.ru/jarticles.html?id=1371> Дата обращения: 19.11.2019
69. URL: rosleshoz.gov.ru
70. Сведения о государственных информационных системах Федерального Агентства Лесного Хозяйства 2019 год // Открытые Данные на сайте Рослесхоза - URL: <http://rosleshoz.gov.ru/opendata/7705598840-InformationSystems> Дата обращения: 17.02.2020
71. Площадь лесных пожаров // URL: http://rosleshoz.gov.ru/rates/forest_fires/area Дата обращения: 17.02.2020
72. Gourlay, C., Leslie, T., Martino, M. (19 February 2020). From a single lightning strike to Australia's largest bushfire. // ABC News. URL: <https://www.abc.net.au/news/2020-02-19/australia-bushfires-how-heat-and-drought-created-a-tinderbox/11976134?nw=0> Дата обращения: 15.04.2020
73. Amazon Forest Fires Increased By 30 % In 2019: Report (references to INPE) // URL: <https://www.ndtv.com/world-news/amazon-forest-fires-increased-by-30-in-2019-report-2161048> Дата обращения: 15.04.2020

74. Кротов, Н. Презентация заместителя руководителя Рослесхоза на форуме "Чистая страна" - 9.12.2019. // URL: <https://ria.ru/20191217/1562490197.html> Дата обращения: 15.04.2020
75. Оценка пожаров по космическим снимкам / Сайт Компании Иннотер // URL: <https://innoter.com/articles/otsenka-pozharov-po-kosmicheskim-snimkam/> Дата обращения: 18.11.2019
76. Замолодчиков, Д.Г. Системы оценки прогноза запасов углерода в лесных экосистемах // Устойчивое лесопользование. – 2011. – № 4 (29). – С. 15-22.
77. Лесной План Республики Карелия на 2019–2028 годы / Утвержден 24 декабря 2018 г. № 731-р. // URL: <https://minprirody.karelia.ru/lesnoe-hozjajstvo/lesnoj-plan-respubliki-karelija-2019-2028/> Дата обращения: 20.02.2020
78. Лесной План Архангельской области РФ на 2019–2028 годы / Министерство природных ресурсов и лесопромышленного комплекса Архангельской области 2018. // URL: <https://yadi.sk/i/644OpWpLehILOA> Дата обращения: 20.02.2020
79. Лесной План Ленинградской области на 2019–2028 годы / ФИЛИАЛ ФГБУ «Рослесинфорг» «Севзаплеспроект». – СПб., 2018. Дата обращения: 20.02.2020
80. Приказ Министерства природных ресурсов и экологии РФ 13.09.2016 г. № 474 «Об утверждении Правил заготовки древесины и особенностей заготовки древесины в лесничествах, лесопарках, указанных в статье 23 Лесного кодекса Российской Федерации». Дата обращения: 05.03.2020
81. Рослесхоз представит в 2020 году портал о лесных ресурсах страны // сайт РИА Новости - URL: <https://ria.ru/20191223/1562723917.html> Дата обращения: 06.03.2020
82. Колесникова, О.Н. Использование ДДЗЗ с КА RapidEye и WorldView-2 для решения задач сельского и лесного хозяйства, компания «Совзонд»
83. Подмосковский опыт учтут при разработке федеральной платформы контроля лесной отрасли // URL: <http://www.gisa.ru/127561.html> Дата обращения: 06.03.2020
84. В Калужской области создали интерактивную карту лесов // URL: <https://pressa40.ru/v-kaluzhskoy-oblasti-sozdali-interaktivnuyu-kartu-mestnyh-lesov/> Дата обращения: 06.03.2020
85. Объемы работ по инвентаризации фонда лесовосстановления в 2020 году увеличатся // Все о российских лесах-URL: http://www.forest.ru/news/forest_restoration/the_amount_of_work_on_inventory_of_reforestation_in_2020/ Дата обращения: 06.03.2020
86. Грабовский, В. И., Замолодчиков, Д. Г. Модели оценки запасов валежа по данным учетов на трансектах // Лесоведение. – 2012. – №. 2. – С. 66-73.

87. Креснов В. В., Страхов В. В., Филипчук А. Н. Национальная инвентаризация лесов в зарубежных странах //Лесохозяйств. информ. – 2008. – №. 10-11. – С. 53-88.
88. Beaudoin A. et al. Mapping attributes of Canada's forests at moderate resolution through k NN and MODIS imagery //Canadian Journal of Forest Research. – 2014. – Т. 44. – №. 5. – С. 521-532.
89. Chirici, G. & Mura, M. & McInerney, D. & Py, N. & Tomppo, E. & Waser, L. & Travaglini, D. & Mcroberts, R. (2016). A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. // Remote Sensing of Environment. 176. 282–294.
90. Stojanova, D. et al. Learning to predict forest fires with different data mining techniques //Conference on data mining and data warehouses (SiKDD 2006), Ljubljana, Slovenia. – 2006. – С. 255-258.
91. Shang, X., & Chisholm, L.A. (2014). Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms. IEEE // Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 7, 2481-2489.
92. Kangas A. et al. Remote sensing and forest inventories in Nordic countries—roadmap for the future //Scandinavian Journal of Forest Research. – 2018. – Т. 33. – №. 4. – С. 397-412.
93. OpenTopography High-Resolution Topography Data and Tools // URL:<https://opentopography.org/> Дата обращения: 12.03.2020
94. USGS EarthExplorer //URL:<https://earthexplorer.usgs.gov/> [United States Inter-agency Elevation Inventory (USIEI)]//URL:<https://coast.noaa.gov/inventory/> Дата обращения: 12.03.2020
95. United States Inter-agency Elevation Inventory (USIEI)//URL:<https://coast.noaa.gov/inventory/> Дата обращения: 12.03.2020
96. LiDAR Cloud Dielmo // URL:<https://www.dielmo.com/en/lidar-cloud/> Дата обращения: 12.03.2020
97. Савченко, А.А., Выводцев, Н.В. Оценка возможностей применения данных дистанционного зондирования при мониторинге санитарного и лесопатологического состояния лесов // Ученые заметки ТОГУ. – 2015. – Т. 6. No 4. – С. 658–661.
98. Терехов, А. Г., Макаренко, Н. Г., Пак, И. Т. Автоматический алгоритм классификации снимков QuickBird в задаче оценки полноты леса. // Компьютерная оптика. – 2014. – Т. 38. – №. 3.

99. Крылов, А.М., Малахова, Е.Г., Владимирова, Н.А. Выявление и оценка площадей катастрофических ветровалов 2009–2010 гг. по данным космической съемки // Изв. Санкт-Петербургской лесотехнической академии. – 2012. Вып. 200. – С. 197–208.
100. McLaren J. R., Turkington R. Ecosystem properties determined by plant functional group identity // Journal of Ecology. – 2010. – Т. 98. – №. 2. – С. 459-469.
101. Wulder, M. A., Franklin, S. E. Understanding forest disturbance and spatial pattern: remote sensing and GIS approaches. – CRC Press, 2006.
102. Северо-Западный ФО: Об Округе // сайт ОО «Деловая Россия» - URL: <https://deloros.ru/severo-zapadnyj-fo.html> Дата обращения: 19.03.2020
103. Прогноз долгосрочного социально-экономического развития Российской Федерации на период до 2030 года // URL: http://www.consultant.ru/document/cons_doc_LAW_144190/ Дата обращения: 19.03.2020
104. Приказ Минприроды РФ N 367 "Об утверждении Перечня лесорастительных зон Российской Федерации и Перечня лесных районов Российской Федерации" (18 августа 2014 г.) // URL: http://www.consultant.ru/document/cons_doc_LAW_169590/ Дата обращения: 20.03.2020
105. Северо-Западный ФО // URL: http://www.priroda.ru/regions/forest/index.php?FO_ID=600&SECTION_ID=586 Дата обращения: 19.03.2020
106. Сведения о землях лесного фонда // URL: <http://rosleshoz.gov.ru/opendata/7705598840-ForestFund> Дата обращения: 19.05.2020
107. Карта арендаторов лесного фонда Карелии / Общественная природоохранная организация «СПОК» // URL: <http://spok-karelia.ru/?id=683> Дата обращения: 20.04.2020
108. URL: <https://forest.kosmosnimki.ru/> Дата обращения: 20.04.2020
109. Демо-версия интерактивной карты ЛПК – URL: <https://lesprominform.ru/maps> Дата обращения: 20.04.2020
110. URL: https://sbo-paper.ru/news/archive_rus/10338 Дата обращения: 19.03.2020
111. Коревко, С. Г. Классификация рисков в лесопромышленном комплексе // Актуальные вопросы экономических наук. – 2008. – № 3-2.
112. Расходы на обеспечение охраны, защиты, воспроизводства лесов на землях лесного фонда (тыс. руб.) // URL: <http://rosleshoz.gov.ru/opendata/7705598840-ForestProtectionReforestationCosts> Дата обращения: 4.04.2020
113. Расходы на лесовосстановление (тыс. руб.) // URL: <http://rosleshoz.gov.ru/opendata/7705598840-ReforestationCosts> Дата обращения: 4.04.2020
114. Информация о результатах плановых и внеплановых проверок, проведенных Федеральным агентством лесного хозяйства и его территориальными органами в пределах

своих полномочий (11 декабря 2019 г.) // URL: <http://rosleshoz.gov.ru/opendata/7705598840-RegisterProvOrganization> Дата обращения: 18.03.2020

115. Концепция интенсивного использования и воспроизводства лесов. – СПб. : ФБУ «СПбНИИЛХ», 2015. – 16 с.

116. Проблемы и перспективы внедрения интенсивного, экономически эффективного, экологически устойчивого и социально ответственного лесного хозяйства // ЛесПромИнформ 2015. – № 2 (108) - URL: <https://lesprominform.ru/jarticles.html?id=4000> Дата обращения: 15.02.2020

117. URL: http://old.gov.karelia.ru/gov/News/2018/12/1220_20.html Дата обращения: 19.03.2020

118. Примеры отечественного опыта устойчивого лесопользования и лесопользования : сб. статей / под общ. ред. Н. Шматкова; Всемирный фонд дикой природы (WWF). – М. : WWF России, 2013. – 240 с.

119. Элбакидзе, М. Модельные Леса Северо-Запада Российской Федерации: взгляд со стороны // Устойчивое лесопользование. 2008. № 1 (17).

120. «Псковский модельный лес». Проект закончен, проект продолжается... // ЛесПромИнформ. – 2009. – № 3 (61). // URL: <https://lesprominform.ru/jarticles.html> Дата обращения: 18.03.2020

121. Шматков, Н.М. "Скандинавская" Модель Интенсивного Лесного Хозяйства // Примеры зарубежного опыта устойчивого лесопользования и лесопользования. – М. : Изд.: Всемирный фонд природы, 2012.

122. Россия лидирует по площади лесов с ответственным лесопользованием. // ЛесПромИнформ. – 2020. – № 2 (148). // URL: <https://lesprominform.ru/jarticles.html?id=5647> Дата обращения: 21.03.2020

123. Каткова, Т. Е. Развитие теории и методов риск-менеджмента в лесном хозяйстве // Экономические науки. – 2011. – № 7. – С. 68-72.

124. Hanewinkel, M., Hummel, S., & Albrecht, A. (2011). Assessing natural hazards in forestry for risk management: a review. // European Journal of Forest Research, 130 № 3. – P. 329-351.

125. Schelhaas M. J., Nabuurs G. J., Schuck A. (2003). Natural disturbances in the European forests in the 19th and 20th centuries // Global Change Biology. — Т. 9. – №. 11. – P. 1620-1633.

126. Налдеев, Д. Ф. Вспышка массового размножения короеда-типографа в национальном парке «Водлозерский» Республики Карелия // Вестник Московского государственного университета леса. // Лесной вестник. – 2009. – № 5. – С. 126-127.

127. Крылов А. М., Соболев А. А., Владимирова Н. А. Выявление очагов короеда-типографа в Московской области с использованием снимков Landsat //Лесной вестник/Forestry bulletin. – 2011. – №. 4.
128. Hanewinkel, M. et al. (2008). Seventy-seven years of natural disturbances in a mountain forest area—the influence of storm, snow, and insect damage analysed with a long-term time series // Canadian Journal of Forest Research. — V. 38. № 8. – P. 2249-2261.
129. Законодательные основы для интенсивного лесного хозяйства // А. Марковский, А. Родионов, ООО «Лесная территория» - URL: <http://www.forest-karelia.ru/wordpress/законодательные-основы-для-интенсив/> Дата обращения: 15.02.2020
130. URL: <http://oldsite.zapoved.ru/activities/> Дата обращения: 15.09.2019
131. Приказ Минприроды России от 05.04.2017. N 156 "Об утверждении Порядка осуществления государственного лесопатологического мониторинга" (Зарегистрировано в Минюсте России 30.06.2017 N 47257). VIII. Дистанционные наблюдения за санитарным и лесопатологическим состоянием лесов.
132. Бореальной лесной платформе 5 лет// ЛесПромИнформ. 2019. № 8 (146). – URL: https://wwf.ru/upload/iblock/6d9/Borealnoi_-lesnoi_-Platforme-5-let.pdf Дата обращения: 14.03.2020
133. Карпачев, А.П. Опыт классификации космоснимка Sentinel- 2a с помощью Semi-Automatic Classification Plugin в QGIS // URL: <https://gis-lab.info/qa/qgis-sacp-sentinel2a.html> Дата обращения: 1.03.2020
134. URL: http://www.mnr.gov.ru/docs/dokumenty_po_voprosam_oopt/akty_lesopatologicheskogo_obsledovaniya_lesnykh_nasazhdeniy/ Дата обращения: март 2020
135. URL: <http://www.forestforum.ru/viewtopic.php?t=20557> Дата обращения: 18.03.2020
136. Леса высокой природной ценности Республики Карелия: Геопортал. // URL: <http://www.hcvf.ru/ru/maps/karelia> Дата обращения: 3.03.2020
137. URL: <https://yadi.sk/d/sHnAB2ys3PYgBe/> Дата обращения: 3.03.2020
138. Гусев, Н. Н., Заварзин, В. В. Лесоустройство : учеб. – М., 2004.
139. Приказ Об утверждении порядка проведения лесопатологических обследований и формы акта лесопатологического обследования // URL:<http://docs.cntd.ru/document/420377904> Дата обращения: 15.03.2020
140. USGS // URL: <http://earthexplorer.usgs.gov> Дата обращения: 8.03.2020
141. Sentinels Scientific Data Hub // URL: <https://scihub.copernicus.eu/dhusm> Дата обращения: 8.03.2020

142. Стыценко Ф. В. Разработка и применение методики и автоматизированной технологии оценки пирогенной гибели лесов на основе спутниковых данных : дис. – (спец. 25.00. 34).–М., 2016.–131 с, 2016.

143. Шихов А. Н. Исследование последствий сильных шквалов и смерчей в Пермском крае с применением данных дистанционного зондирования Земли //Географический вестник. – 2013. – №. 1 (24).

144. Князева, С.В., Королева, Н.В., Эйдлина, С.П. Динамика лесовосстановления на территории очага сибирского шелкопряда по материалам съемок высокого и сверхвысокого разрешения // III Всероссийская научная конференция: Научные основы устойчивого управления лесами. 30 октября – 1 ноября 2018 г. – М., 2018.

Приложение. Модель Random Forest

(по национальному парку "Паанаярви", Карелия)

```
import numpy as np
from sklearn import cluster
from osgeo import gdal, gdal_array
import pandas as pd
from pandas import set_option
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
```

```
# Telling GDAL to throw Python exceptions, and register all drivers
gdal.UseExceptions()
gdal.AllRegister()
```

Загрузка файлов обучающей выборки - целевой признак

```
# Путь к директории с файлами
datadir = '/content/drive/My Drive/Colab Notebooks/PaanRF/'
```

```
# Создаем список с полными именами файлов (итерируя через него с помощью функции read_tiff создадим вектора)
datadir = '/content/drive/My Drive/Colab Notebooks/PaanRF/'
n_files = 5 # количество классов в классификационной модели и файлов
names = ["water", "infrastructure", "tree1", "tree2", "damage"]
filename_lst = [datadir + 'water.tif', datadir + 'infrastr.tif', datadir + 'tree1.tif', datadir + 'tree2.tif', datadir + 'damage.t
```

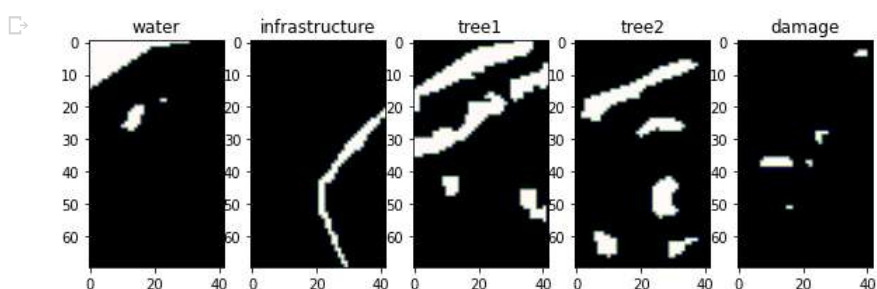
```
# Функция чтения файла tiff из внешней среды, и преобразования ее в вектор с размерами (width*height, 1)
def read_tiff(filename, band_to_read):
    # Запрашиваем tiff из google-drive.
    ds_file = gdal.Open(filename, gdal.GA_ReadOnly) # ds - DataSet-object
    # В изображениях обучающей выборки только 1 канал (1 и единственный), но функция м.б. трансферабельна на многоканальный композит
    ds_band = ds_file.GetRasterBand(band_to_read)
    # Преобразуем в 2D массив
    ds_arr = ds_band.ReadAsArray()
    # Преобразуем в 1D вектор
    ds_vec = ds_arr.reshape((-1,1))

    return ds_vec
```

```
# Создадим список векторов обучающей выборки
ds_list = []
for path in filename_lst:
    ds_list.append(read_tiff(path, 1))
```

```
fig, axs = plt.subplots(1, 5, figsize=(10, 3), facecolor='w', edgecolor='k')
axs = axs.ravel()
for i in range(5):
    axs[i].imshow(ds_list[i].reshape((70,42)), cmap="gist_earth")
    axs[i].set_title(names[i])
```

```
plt.show()
```



```
# Создаем dataframe где в строках - экземпляры (пиксели), в колонках признаки (значения каналов, индексы, etc.)
df = pd.DataFrame()
pixels = [pix for pix in range(len(ds_list[0]))]

# Добавляем данные в dataframe
df['pix'] = pixels
df['water'] = ds_list[0]
df['infra'] = ds_list[1]
df['tree1'] = ds_list[2]
df['tree2'] = ds_list[3]
df['damage'] = ds_list[4]
# df.head(10)
```

▼ Слияние классов в 1 целевой признак 'class'

```
# Создадим один целевой признак из 5 классов
df['class'] = 0 # по умолчанию значение параметра 'class' 0
df.loc[(df['water'] == 1), 'class'] = 1 # water
df.loc[(df['infra'] == 2), 'class'] = 2 # infra
df.loc[(df['tree1'] == 3), 'class'] = 3 # fir - tree1
df.loc[(df['tree2'] == 4), 'class'] = 4 # mixed - tree2
df.loc[(df['damage'] == 5), 'class'] = 5 # damage
df.head(3)
```

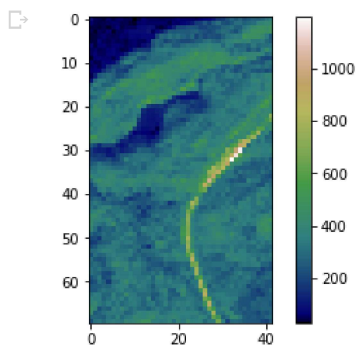
```
pix  water  infra  tree1  tree2  damage  class
0      0     1.0    0.0    0.0    0.0    0.0     1
1      1     1.0    0.0    0.0    0.0    0.0     1
2      2     1.0    0.0    0.0    0.0    0.0     1
```

▼ Загрузка мультиспектрального изображения (поканально)

```
# Считываем выдел в комбинации 2,3,4,8,11. Разрешение 20 м.
# Sentinel-2 data is DOS atmospherically corrected. Pixels store BOA reflectance values.
image_path = datadir + 'area.tif'
img_ds = gdal.Open(image_path, gdal.GA_ReadOnly)
img_ds_bnds = []
band_nums = [2,3,4,8,11]
for band_num in range(1, 6):
    img_ds_bnds.append(read_tiff(image_path, band_num))
len(img_ds_bnds)
```

```
5
```

```
plt.figure()
plt.imshow(img_ds_bnds[1].reshape((70,42)), cmap="gist_earth")
plt.colorbar()
plt.show()
```



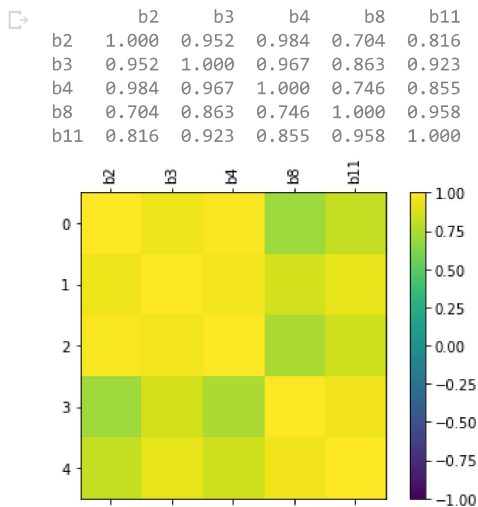
```
# Добавляем данные каналов снимка в dataframe
for i in range(len(img_ds_bnds)):
    band_name = 'b' + str(band_nums[i])
    df[band_name] = img_ds_bnds[i]
features = ['b2', 'b3', 'b4', 'b8', 'b11']
df.shape
```

↳ (2940, 12)

▼ Корреляция признаков

признаки = значения в каналах

```
dfBands = df[features]
# Корреляционная матрица из коэффициентов Пирсона
set_option('display.width', 150)
set_option('precision', 3)
bandCorr = dfBands.corr(method='pearson')
print(bandCorr)
# Plotting
corFig = plt.figure()
figAxis = corFig.add_subplot(111) # 111 means a 1 x 1 grid for the 1st subplot
corAx = figAxis.matshow(bandCorr, vmin=-1, vmax=1)
corFig.colorbar(corAx)
indices = [0,1,2,3,4]
varr = features
plt.xticks(indices, varr, rotation=90)
plt.show()
```



▼ Масштабирование признаков

```
# Функция масштабирования параметров
minmaxScaler = preprocessing.MinMaxScaler()
def scale_band(df_to_scale, bnd):
    # bnd: string
    xbnd = df_to_scale[[bnd]].values.astype(float)
    bndn = bnd + 's'
    df_to_scale[bndn] = minmaxScaler.fit_transform(xbnd)
    # чтобы избежать нулевых значений добавим маленькое число i.e. 0.00001
    df_to_scale[bndn] = df_to_scale[bndn] + 0.00001

# Масштабируем данные каналов снимка в dataframe
for i in range(len(img_ds_bnds)):
    band_name = 'b' + str(band_nums[i])
    scale_band(df, band_name)
features_scaled = ['b2s', 'b3s', 'b4s', 'b8s', 'b11s']
```

▼ Стандартизирование признаков

```
x_df = df.loc[:, features_scaled].values
x_df = StandardScaler().fit_transform(x_df)
features_stand = ['b2st', 'b3st', 'b4st', 'b8st', 'b11st']
stand_df = pd.DataFrame(data = x_df, columns = features_stand)
stand_df = pd.concat([stand_df, df[['class']]], axis = 1)
stand_df
```

↳

	b2st	b3st	b4st	b8st	b11st	class
0	-1.609	-1.839	-1.465	-1.944	-1.672	1
1	-1.765	-2.166	-1.750	-2.676	-2.200	1
2	-1.713	-2.206	-1.809	-2.702	-2.419	1
3	-1.622	-2.280	-1.734	-2.651	-2.454	1
4	-1.557	-2.149	-1.801	-2.722	-2.487	1
...
2935	-0.724	-0.614	-0.717	-0.225	-0.613	0
2936	-0.672	-0.671	-0.793	-0.347	-0.718	0
2937	-0.841	-0.940	-0.927	-0.278	-0.740	0
2938	-1.128	-1.022	-0.986	-0.426	-0.751	0

▼ PCA

```
# https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60
from sklearn.decomposition import PCA
pca = PCA(n_components=3)
principalComponents = pca.fit_transform(x_df)
principalDf = pd.DataFrame(data = principalComponents, columns = ['pc1', 'pc2', 'pc3'])
principalDf.shape
```

```
↳ (2940, 3)
```

▼ Создание полного набора данных модели

```
full_df = pd.concat([principalDf, df[['class']]], axis = 1)
model_df = full_df[full_df['class'] != 0]
model_df = model_df.reset_index()
del model_df['index']
model_df
```

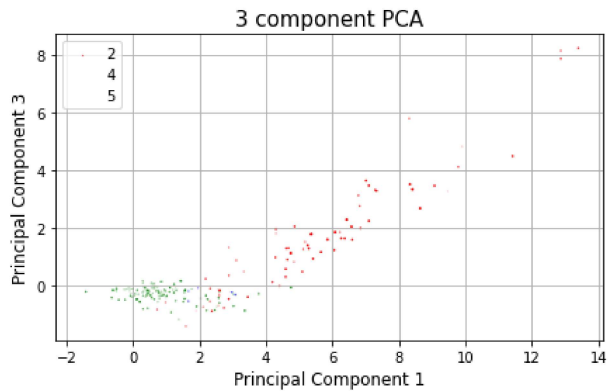
```
↳
```

	pc1	pc2	pc3	class
0	-3.807	0.338	-0.255	1
1	-4.706	0.795	-0.286	1
2	-4.838	0.895	-0.136	1
3	-4.792	0.943	-0.128	1
4	-4.777	1.021	-0.059	1
...
1290	1.013	-0.336	0.316	4
1291	6.427	2.295	0.819	2
1292	4.296	1.976	0.585	2
1293	2.607	-0.146	0.446	2
1294	4.746	1.139	0.745	2

1295 rows × 4 columns

```
fig = plt.figure(figsize = (7,4))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel('Principal Component 1', fontsize = 12)
ax.set_ylabel('Principal Component 2', fontsize = 12)
ax.set_ylabel('Principal Component 3', fontsize = 12)
ax.set_title('3 component PCA', fontsize = 15)
targets = [2, 4, 5]
colors = ['r', 'g', 'b']
for target, color in zip(targets,colors):
    indicesToKeep = model_df['class'] == target
    ax.scatter(model_df.loc[indicesToKeep, 'pc1'],model_df.loc[indicesToKeep, 'pc2'],
               model_df.loc[indicesToKeep, 'pc3'], c = color)
ax.legend(targets)
ax.grid()
```

```
↳ /usr/local/lib/python3.6/dist-packages/matplotlib/collections.py:886: RuntimeWarning: invalid value encountered in sqrt
scale = np.sqrt(self._sizes) * dpi / 72.0 * self._factor
```



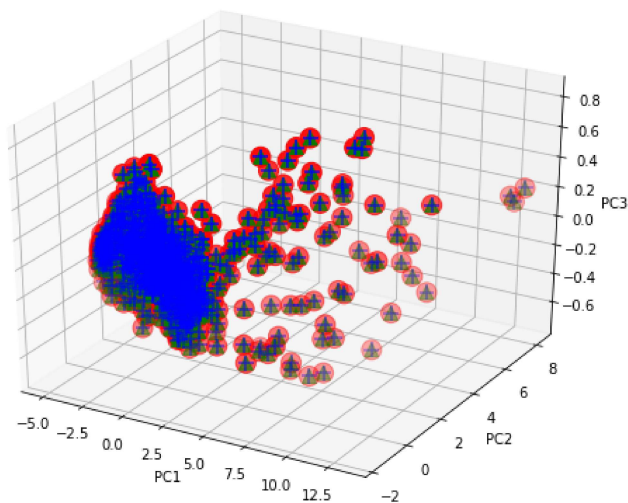
```
# https://matplotlib.org/2.1.1/gallery/mplot3d/scatter3d.html
from mpl_toolkits.mplot3d import Axes3D
```

```
fig = plt.figure(figsize=(9,7))
ax = fig.add_subplot(111, projection='3d')
# для каждой компоненты свой стиль и цвет и размер
for c, m, s in [('r', 'o', 200), ('g', '^', 100), ('b', '+', 120)]:
    xs = np.array(model_df['pc1'])
    ys = np.array(model_df['pc2'])
    zs = np.array(model_df['pc3'])
    ax.scatter(xs, ys, zs, c=c, marker=m, s=s)

ax.set_xlabel('PC1')
ax.set_ylabel('PC2')
ax.set_zlabel('PC3')

plt.show()
```

↳



▼ Инициализация и обучение модели RF

- ▮ разделение на Train & Test
- ▮ инициализация модели и обучение на Train-set
- ▮ тестирование на Test-set

```
# Извлекаем целевую переменную/признак в вектор y
y = model_df.pop('class')
# Разделяем данные на Training и Testing поднаборы
X_train, X_test, y_train, y_test = train_test_split(model_df, y, test_size = 0.4, random_state=188)
```

```
from sklearn.ensemble import RandomForestClassifier
# Создаем модель классификатора, инициализирую RF object с параметрами
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
```

▼ Сохранение результата

```
# Сохраняем результат классификации как geotiff изображение
ds = gdal.Open(datadir+'area.tif')
band = ds.GetRasterBand(4)
arr = band.ReadAsArray()
[cols, rows] = arr.shape
format = "GTiff"
driver = gdal.GetDriverByName(format)

fname = datadir + 'RF_paan_2_89.tif'
outClassifiedRaster = driver.Create(fname, rows, cols, 1, gdal.GDT_Byte)
outClassifiedRaster.SetGeoTransform(ds.GetGeoTransform()) #устанавливаем аналогичный исходному тип геом преобразования
outClassifiedRaster.SetProjection(ds.GetProjection()) # и аналогичную проекцию

outClassifiedRaster.GetRasterBand(1).WriteArray(predsRe)

outClassifiedRaster.FlushCache() # удаляем из памяти
del outClassifiedRaster # удаляем данные, не geotiff изображение
```

```
# Обучаем модель
rf_model.fit(X_train, y_train)

RandomForestClassifier(bootstrap=True, ccp_alpha=0.0, class_weight=None,
                        criterion='gini', max_depth=None, max_features='auto',
                        max_leaf_nodes=None, max_samples=None,
                        min_impurity_decrease=0.0, min_impurity_split=None,
                        min_samples_leaf=1, min_samples_split=2,
                        min_weight_fraction_leaf=0.0, n_estimators=100,
                        n_jobs=None, oob_score=False, random_state=42, verbose=0,
                        warm_start=False)

from sklearn.metrics import accuracy_score
# Предсказываем классы training-набора
train_preds = rf_model.predict(X_train)

# Точность модели на тренировочной выборке
train_acc = accuracy_score(y_train, train_preds)
print(train_acc)

1.0

# Тестирование модели
test_preds = rf_model.predict(X_test)
# Точность модели на тренировочной выборке
test_acc = accuracy_score(y_test, test_preds)
print("Точность модели accuracy = {:.f}".format(test_acc))

Точность модели accuracy = 0.808880
```

▼ Применение модели

```
full_preds = rf_model.predict(full_df.drop('class', axis=1))

predsRe = np.reshape(full_preds, (70,42))
print (predsRe.shape)

plt.figure(figsize=(10,9))
plt.imshow(predsRe, cmap="gist_earth")
plt.colorbar()
plt.show()
```

